



**AUSTRALIAN NATIONAL DATA SERVICE
(ANDS) BUSINESS PLAN, 2009/10**

AUSTRALIAN NATIONAL DATA SERVICE (ANDS) BUSINESS PLAN, 2009/10

1	Introduction	4
2	ANDS Context	5
2.1	ANDS Objectives.....	5
2.2	ANDS Principles.....	6
2.3	ANDS Scope	7
2.4	Discovery and the Australian Research Data Commons.....	8
3	Status of Project	10
4	Research Infrastructure	12
4.1	Developing Frameworks	12
4.2	Providing Utilities.....	16
4.3	Seeding the Commons.....	18
4.4	Building Capabilities	21
5	Program Engagement Strategies	24
5.1	Program Participants.....	25
5.2	Targeted Program Engagement Partner Summary	26
5.3	Engagement with Government about data	26
5.4	Other forms of engagement	26
6	Governance	27
6.1	Steering Committee.....	27
6.2	Management structure and responsibilities	27
7	Promotion.....	32
8	Access and Pricing	33
8.1	Access and Pricing.....	33
8.2	Level of usage	34
9	Financial and Human Resources	35
10	Milestones	38
10.1	Milestones for 2009-10.....	38
10.2	Key performance indicators	39
11	NeAT Projects.....	42
11.1	NeAT Status	42

11.2 ARCS-ANDS Agreed NeAT Governance:43

11.3 Spatial Information Services Stack (SISS).....44

11.4 Marine and Climate Data Discovery and Access Project (MACDDAP).....48

11.5 Data Integration and Annotation Services in Biodiversity (DIAS-B)51

11.6 Collaborative Integration and Annotation Services for Australian Literature Communities (Aus-e-Lit) Project.....56

11.7 ASSDA Services for e-Social-Science (ASeSS).....59

11.8 A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities (DataMINX).....62

11.9 NeAT Round 2 Ideas that are being developed as possible proposals.....67

1 Introduction

After extensive planning through consultations that created *Towards the Australian Data Commons* and a comprehensive establishment project, the Australian National Data Service (ANDS) was established effective January 1st, 2009. Many parties recognise the importance of data, and they are engaging with ANDS – there was clearly pent-up demand for engaging with data issues. Many researchers and institutes have begun engaging with ANDS on their data issues, and the challenge for ANDS is to respond to this demand and support researchers and institutions in using their data more effectively.

In this document plans for 2009-10 are described that both meet the needs that have been expressed to date but also shape Australia's ability to use its research data effectively. Consequently the plan reflects a balance between the response to the groups who are "ready, willing and able" to engage now, whilst maintaining a strategic focus on long-term cultural, procedural and technical change towards the increasing centrality of data in research.

In the first six months ANDS has responded to many individual requests for engagement, and have initiated many others. It is clear that ANDS could expend all of its effort by responding to requests, but, as described in the 2008-09 Business Plan, ANDS has devoted part of its effort in this manner, particularly in the "*Seeding the Commons*" program, but have also devoted effort to developing national services, national responses to the *Code for the Responsible Conduct of Research* and advice that will enable many parties to better manage their research data. By March 1, 2009, ANDS had 16 staff (including the Executive Director and three Deputy Directors). Work had commenced on all of the four programs, and ANDS was actively engaging with its stakeholders. A workshop on "Supporting the Data Lifecycle" was jointly sponsored and held by ANDS and ARCS in Sydney in February 2009. An ANDS Collections Registry and a Persistent Identifier prototype have been created and are due to be commissioned as services by July 2009. There is now a feed from crystallography data sets into the ANDS Collections Registry that describes persistently-identified data that have been made available and are referred to in publications. NeAT Round 1 projects are underway and a second round of NeAT projects has been identified and are described later in this document.

This Business Plan does not deviate from the 2008-09 Business Plan in any substantial way – the principal change is that there is an increased emphasis in engaging with our partners from a whole of ANDS perspective, rather than selecting targets for Seeding activities, and others for Capabilities, for example. It is clear in the engagements to date that the partners are interested in more general engagement – this has the consequence that ANDS needs to have a greater emphasis on "customer relationship management" so that partners do not have to navigate through parts of ANDS; rather ANDS will ensure that teams will include those parts of ANDS that are relevant to their need. In particular many engagements are envisaged where ANDS has expertise from both Seeding the Commons and Building Capabilities represented in field teams that will have data management and data technology expertise.

By the end of 2009-10 ANDS expect that there will be many researchers who have published their data using ANDS persistent identifiers into a data store that feeds information to the ANDS Collections Registry. ANDS expects that researchers will be able to find a wide variety of data sets using the ANDS data pages through a variety of discovery paths, and that more institutions will be successfully engaged in meeting their responsibilities described in the "Code for the Responsible Conduct of Research". Most importantly ANDS will have engaged the research community sufficiently that researchers stop asking, "why should I publish my research data?" but are asked and asking, "why shouldn't I publish my research data?"

2 ANDS Context

2.1 ANDS Objectives

The document entitled *Towards the Australian Data Commons*¹ (TADC) succinctly summarises the high level vision for ANDS in its overview:

The development of ANDS is intended to provide the essential meeting place where the Australian path forward for research data management can evolve and where a vision can be achieved. This vision will articulate over time policies and guidelines that are readily understood and interpreted while simultaneously creating exemplars of best practice covering:

- research data ownership and the roles and responsibilities associated with ownership;
- access to research data collected and maintained with public funding; and
- best practice for the curation of experimental, research and published data. (p. 3)

In support of this, *Towards the Australian Data Commons* identifies 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 identifies 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)

To deliver against these objectives, ANDS has four inter-related programs of activity (Developing Frameworks, Providing Utilities, Seeding the Commons, Building Capabilities). ANDS also funds specific development activity towards the aims of the Providing Utilities

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

and Seeding the Commons programs under the banner of the National e-Research Architecture Taskforce (NeAT). Please see section 11 for details of these NeAT Projects.

Working with the whole sector, by the end of three years the ANDS project will have made progress towards delivering this vision in a number of different ways. ANDS will have established national data discovery services that enable authenticated access to research data across the Australian Research Data Commons. It will be possible to refer to the data using referencing that is as stable as bibliographic references. There will be substantially improved data availability through better thought out data management. This will enable researchers to create data, store it in institutional repositories, and make it accessible more widely by using the most appropriate national and international standards.

The data environment will also have changed – institutions will be expected to have and support data management plans, and any researcher seeking support through a number of government funding agencies will be expected to describe how the data generated through the project will be managed throughout its lifecycle. Progress towards these outcomes will not be uniform or comprehensive after three years, but all the stakeholders will be aware that change is needed and underway.

As a result, at the end of the funding period there will be a number of well-publicised stories of high value data being used beyond its initial use, concretely demonstrating to researchers the value of the Australian Research Data Commons.

2.2 ANDS Principles

In responding to the objectives and program requirements, ANDS continues to follow these principles:

2.2.1 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.

2.2.2 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.

2.2.3 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.

2.2.4 Service Provision

ANDS is focussed on service provision, not research and exploration; its programs will develop, integrate, and continually improve production-level systems in support of well-understood services. NeAT will fund the development of more innovative and exploratory domain-focused initiatives that may become ANDS services in later years.

2.2.5 Strategic Partners

ANDS recognises the need to actively engage with external institutions relevant to the Australian Research Data Commons, including the Australian Access Federation (AAF) and the Australian Research Collaboration Service (ARCS).

2.2.6 Stores

ANDS assumes an environment where storage and long-term curation occur in 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.

2.2.7 Sustainability

Research data management requires a long-term commitment. ANDS has developed its three-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.

2.3 ANDS Scope

2.3.1 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), Departments of Primary Industry
- 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 National Collaborative Research Infrastructure Strategy (NCRIS) capabilities. This means that the specific requirements of the Humanities and Social Sciences will need to be taken into account.

2.3.2 ANDS Community

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

2.3.3 Data

ANDS is concerned with data that is produced by researchers as well as data that is used by and made accessible to them. 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.

2.4 Discovery and the Australian Research Data Commons

Discovering research data collected and maintained with public funding is one of the core aims of ANDS as articulated in *Towards the Australian Data Commons*. Figure 1 to the right shows an overview of how ANDS plans to facilitate this. ANDS will work to improve deposit of data, discovery of data, access to that data, and the usability of the data once accessed.

There are many forms of access that are important – from access to local data developed by the local research team to discovery of a data set that leads to a major innovation by combining different perspectives. There will be many technologies that may be needed to support this. These include rich map-based interfaces to support seamless exploration of spatial data, free text search over descriptions of collections, and exploring well described ontologies that have been developed by a community.

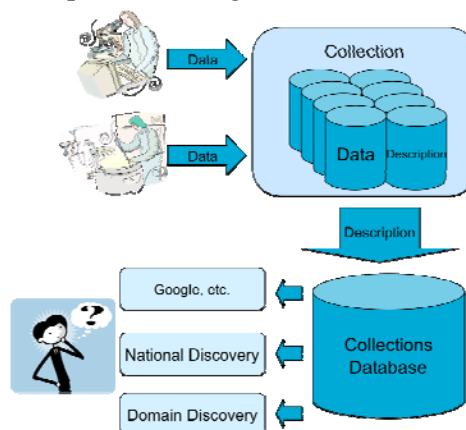


Figure 1: ANDS Discovery and Access Overview

It is a fundamental requirement for ANDS to ensure that researchers be provided with high quality discovery of research data in a form that makes sense to them. To that end, ANDS is committed to build some pieces of the discovery puzzle. Some components have already been built and other pieces again are being built for specific disciplines through other parts of NCRIS, and in some cases through specific NeAT projects.

Sometimes, researchers will gain access directly to the data (most likely when the data is already well understood). In other instances the access may go via information about that data (perhaps described in its metadata, or a relevant document) or to the collection within which the data sits. These latter cases will provide the searcher with a broader perspective on the potential relevance of the data.

The ANDS establishment team conducted a series of consultations to better understand how researchers need to discover and access information about collections of data. One outcome is that this discovery may need to take many forms, and ANDS will need to respond in multiple ways (see figure 1). Some of these include:

- enabling web search using Google etc. by publishing web pages from the registry that manages the collections database
- enabling both free text search and metadata search with a discovery service that sits directly on top of the registry
- providing browsing support with pathways through the collections using such tools as finding aids and other methods.
- creating a search interface that enables domain or problem specific portals such as the Atlas of Living Australia to augment their own searches with searches to the collections registry that provide pointers to collections outside the domain that may be of interest – an equivalent to the “see also” service on Amazon.
- referring directly to domain and institutional search services using domain specific access interfaces from a portal

- providing a portal that supports these discovery services and provides other services on top of the collection database that underlies the registry service

Whilst all of these forms of access are important, ANDS believes that it is very important early in ANDS to get a substantial number of collections described and registered, make the descriptions available for web search, and make the set of collections directly searchable. This will allow a central discovery service to be provided, and also enable “see also” services to be added to domain and institutional discovery services when it is desirable to do so. Activity in support of this approach will need to be carried out in all of the ANDS programs.

3 Status of Project

The Australian National Data Service is one of the components of the Platforms for Collaboration (PfC) capability. 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 the Australian National Data Service (ANDS) Establishment Project. ANDS is part of the Platforms for Collaboration capability within the National Collaborative Research Infrastructure Strategy (NCRIS). This activity was originally funded through to the end of June 2008. Prior to this deadline the Department of Innovation, Industry, Science and Research (DIISR) (which took over responsibility for NCRIS after the change of Government) indicated they wished to extend this deadline to ensure sufficient consultation with a range of stakeholders. At around the same time, it was decided to re-focus the original Repositories and Researcher Practice programs from *Towards the Australian Data Commons* into Seeding the Commons and Building Capabilities. This resulted in the ANDS Establishment Project being extended until 31 December 2008.

To develop a fully-fledged ANDS in anticipation of and building on the outcomes of the ANDS Establishment Project, a Collaboration Agreement was signed between the foundation parties. A new NCRIS Funding Agreement for Implementing the ANDS between DIISR and Monash University was executed on 17 September 2008, with NCRIS funding of \$21 million over 2008-2011.

Both the Draft Interim Business Plan submitted in September 2008, and the Interim Business Plan submitted in December 2008 and the Business Plan for 2008-09 submitted in March 2009 under the new Funding Agreement were predicated on 12 full months of activity. As ANDS proper did not commence until January 1 2009, the activity and outcomes that were anticipated to occur in this first calendar year will not all be completed by the end of 2007/08. Some of the “year 1” activities in the previous Business Plan will thus occur in 2009/10.

By March 1, 2009, ANDS had 16 staff (including the Executive Director and three Deputy Directors). Work had commenced on all of the four programs, and ANDS was actively engaging with its stakeholders. A workshop on *Supporting the Data Lifecycle* was jointly sponsored and held by ANDS and ARCS in Sydney in February 2009. An ANDS Collections Registry and a Persistent Identifier prototype have been created and are due to be commissioned as services by July 2009. There is now a feed from crystallography data sets into the ANDS Collections Registry that describes persistently-identified data that have been made available and are referred to in publications. NeAT Round 1 projects are underway and

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

a second round of NeAT projects have been identified and are described later in this document.

4 Research Infrastructure

For researchers to work in the new 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 repository to store their data well
- 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
- the training and training materials that enable the infrastructure to be used well

As described in the 2008-09 Business Plan, the research infrastructure provided by ANDS to enable all of the above will continue to be delivered through four programs of activity:

- *Developing Frameworks* – the frameworks that will enable research data producing institutions to capture, manage and share research data;
- *Providing Utilities* –services that reduce the cost of capture and ease the task of discovery;
- *Seeding the Commons* – improving local data capture and populating the data commons; and
- *Building Capabilities* - improving Australia’s capability to manage its research data.

These activities will be delivered in a co-ordinated manner that will ensure that partners simply engage with ANDS, not with the individual programs. This has the consequence that we need to have a greater emphasis on “customer relationship management” so that partners do not have to navigate through parts of ANDS, but that ANDS will ensure that teams will include those parts of ANDS that are relevant to their needs. In particular many engagements are envisaged where expertise from both Seeding the Commons and Building Capabilities are represented.

Some elements of the data fabric for the Australian Research Data Commons will also be delivered through the Australian Research Collaboration Service (ARCS). ANDS will work with the National e-Research Architecture Taskforce (NeAT) in defining projects that will primarily contribute towards the Seeding the Commons and Providing Utilities programs. Some of the outputs of these activities will appear in ANDS programs in later years.

4.1 Developing Frameworks

4.1.1 Program Aims

To influence relevant national policies, including undertaking the development of policies where appropriate.

To build a common understanding of data management issues and solutions across government departments (commonwealth and state), research funding agencies and research-intensive organisations.

4.1.2 Program Overview

The Developing Frameworks program aims to influence and simplify the overall policy framework within which the ANDS goal is to be achieved, as well as defining how activities by researchers (in order to comply with their grant conditions) and institutions (in order to comply with funding requirements) can contribute to a national research data commons. This program will progressively work to simplify and reduce the number of licenses under which data is created and shared.

ANDS recognises that there is a number of existing activities in the cross-government data management space; this program will, in part build bridges between these activities and research data management. The focus will, where possible, be on the higher-level aggregations of data-managing entities rather than the individual entities themselves. Recommendations 7.8, 7.9, 7.10 and 7.14 of the Cutler Innovation Review support many of the principles important to ANDS, and so our activities are consistent with these recommendations.

The primary collaborators in the Developing Frameworks Program are:

- 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 the Department of Innovation, Industry, Science and Research (DIISR) and the Department of Health and Aging
- research funding schemes such as the Australian Research Commission (ARC), and the National Health and Medical Research Council (NHMRC)
- discipline leaders within institutions; and
- research office staff at institutions.

4.1.3 Program Themes

Leadership and Advocacy

The Developing Frameworks Program will undertake a policy leadership role across the Australian research community. It will achieve this by establishing a series of Australian Research Data Commons Policy Forums (ANDS Policy Forums) where overarching issues of importance to stakeholders can be addressed; by participating in relevant government forums and by ensuring that data management issues are represented to government enquiries and reviews. It will contribute to national policy development through holding issues-based events resulting in documented policy recommendations and will hold thought-leader seminars and workshops to advance policy understanding. It will work with overseas bodies to encourage joint activities and it will undertake consultancies with particular communities or institutions to assist with developing their capacity.

Liaison

Liaison with state and national governments, national data sharing organisations, research funding agencies and research intensive organisations will be integral to contributing to

national policy development, to creating shared understandings and acting as a bridge between ANDS and those government instrumentalities holding research data collections likely to be of interest to researchers in other sectors. In addition, there will be a need to liaise with appropriate overseas organisations such as the DCC, JISC, SURFNET, CoDATA and the NSF.

Policy and plan development and consultancy

This is fundamental to improving data management and access at both national and institutional levels. The Developing Frameworks program will address the significant need for appropriate institutional policies by developing data management frameworks, models, and guidelines suitable for implementation at research-producing institutions at various levels. These instruments will explicitly address data retention requirements, and the roles and responsibilities of the various stakeholders. The need to improve deposit rates and increase data re-use will be addressed through development of a strategy designed to complement the efforts of government and research funding agencies. It will develop policy guidelines for data curation, re-use and referencing, and review the legal protocols related to data creation and management, identifying gaps where they exist. It should be noted that data management planning and implementation will need to occur at a policy level (overall frameworks), within institutions (alignment of institutional practice with national, international and discipline policies and practices), and within research groups (compliance with discipline practices as well as national and institutional requirements). This policy and plan development will occur with an awareness of the needs of and constraints on researchers, and the differences between disciplinary practices.

Data federation governance and agreements

The creation of data federations is dependent on well-founded governance and mutually acceptable agreements between participants. This program will provide policy and governance support to discipline-specific data federations through the provision of consultancy services and the development of appropriate governance frameworks and agreements.

4.1.4 Proposed Activities for 2009/10

Leadership:

- maintain a strong stakeholder engagement strategy, including establishment of the ANDS Policy Forums; and proactively seeking membership/ involvement in other policy forums;
- review existing state and commonwealth data management and records management policies and guidelines;
- contribute to national policy development through issues-based events resulting in documented policy recommendations;
- prepare submissions and responses to relevant government enquiries, reviews, working groups, etc. (scope dependant on demand); and
- advocacy and outreach for data management policy and planning.

Liaison:

- investigate, establish, and build on connections between ANDS and government instrumentalities that hold data collections relevant to researchers;

- consult with and advise DIISR to ensure that good research data management practice is incorporated into Research Training Scheme (RTS) activities; and
- develop relationships with equivalent activities overseas to share approaches to national and international policy activities that can inform ANDS.

Policy development and implementation:

- develop model data management policies, plans and implementation guidelines (as well as associated strategies) for implementation at research-producing institutions at various levels: institutions, research program leaders, research project managers, individual researchers and post-graduate researchers;
- develop consensus based policy guidelines for data re-use, access rights, and data object referencing, consistent with international and discipline standards; and
- adopt, adapt and embed appropriate licence agreements for data sharing as the preferred option, and support other legal aspects of data creation and management.

4.1.5 End of 2009/10 outcomes

By the end of 2009/10 ANDS will have:

- continued discussions with key public sector agencies on clarifying access to data and some data brought into the Commons (through collaboration with Seeding the Commons);
- developed guidelines on how to comply with the data management requirements in the Australian Code for the Responsible Conduct of Research through an ANDS Policy Forum, and trialled these within an institution;
- conducted further ANDS Policy Forums to discuss, inform and create model policy in specific areas such as data management processes in institutions and data-sharing policies;
- trialled a model institutional data management plan within an institution;
- used work undertaken by the OAK Law project to create user advice on IP issues related to data management
- increased the number of institutions with research data management plans (through collaboration with the Building Capabilities program);
- provided written advice on data management to the Research Training Scheme (through collaboration with the Building Capabilities program); and
- provided governance, policy and other support for the creation of a discipline based data sharing initiative.

4.1.6 End of 2010/11 enduring changes

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

- selected publicly funded data producing organisations provide researchers with improved levels of access to their data;
- the majority of the ANDS constituency will have a research data management plan based on (or consistent with) an agreed model, and which is updated and refined on an annual basis;

- policies and SLAs exist (or at least, gaps in such policies have been identified) to support the creation of new community data federations; and
- research funding agencies are in a position to require good data management and accessibility as a condition of research grants.

4.2 Providing Utilities

4.2.1 Program Aims

To ensure necessary technical and 24x7 operational services are provided so that the content in repositories can be aggregated into federations to underpin the development of a research data commons.

To ensure that services develop and evolve to meet changing data reuse requirements.

4.2.2 Program overview

The Utilities program will provide fundamental utility services for a cohesive network of data collections and will provide discovery, access and other value-added services across the resulting data commons. A technical consultancy will also be available to assist integrating research and government instrumentality repositories and registries with core ANDS utilities.

The primary targets for the Utilities program will be:

- content providers and consumers in the data commons;
- data facility managers and administrators; and
- research communities building and operating (new or already existing) data federations.

4.2.3 Program Themes

Activity in this program will focus on three major theme areas, some of which will be partially fulfilled through NeAT project activity.

Federation Utilities

ANDS provides a range of utility data services at a sector-wide level (such as cross-discipline discovery services; national collection registry; persistent identifier service; federation registry; access policy registry). As appropriate these utility services either aggregate information nationally or provide component services across several NCRIS domain areas and other research communities. This ANDS program will also improve existing services and develop new utility services. Robust utility services are delivered under a service delivery framework that defines the roles and responsibilities of those providing and participating in the services.

The human-facing utility services will use AAF services for authentication once available, and ARCS authorisation services (informed by ANDS requirements) for access control.

Federation Utilities National Framework

ANDS will catalyse and support data federation utilities at the level of research communities and their federations. For example ANDS will work with selected NCRIS domain areas to

assist in the establishment of data federations specific to their domain. ANDS will also work to develop and maintain community consensus within the research and public sector on a technical framework for the interoperability of data utility services (registries, metadata catalogues, schemas and harvesting guidelines, search services, etc).

Technical Consultancy

ANDS offers specialist technical advice and consultancy services around data federation utility services. This advice and assistance applies to communities and organisations seeking to create their own data federations and also to those seeking to make use of ANDS utility services.

4.2.4 Proposed Activities for 2009/10

Federation Utilities:

- commission fundamental national utility services to support:
 - users discovering the existence of some data collections and objects;
 - users (and software) generating and resolving persistent digital identifiers that point to digital objects;
- develop a roadmap for expanded services that enable “publication” of data and assist NCRIS domain areas, research institutions and other research communities to improve the quality, discoverability, intelligibility and usability of research data;
- assist with communication and knowledge transfer about these utility services; and
- develop relationships with equivalent activities overseas to share approaches to national utilities (such as registries) that can inform ANDS.

Federation Utilities National Framework:

- convene (together with the Seeding the Commons program) an Australian Research Data Commons Technical Forum (ANDS Technical Forum);
- engage with discipline and community federations on their data federation requirements; and
- develop and maintain a data federations reference model and service integration framework.

Technical Consultancy

- assist potential users in the uptake of utilities services;
- identify strategic (and willing) data federations for partnership;
- as part of the overall ANDS engagement process, identify and implement technical solutions to assist NCRIS domain areas, institutions and research communities; and
- integration of ANDS registries with international initiatives (NSF and JISC).

4.2.5 End of 2009/10 outcomes

By the end of 2009/10 ANDS will have:

- enabled a number of NCRIS domain areas use of ANDS utility services and/or establish their own. Initial targets include: ABIN, ALA, APN, APF, IMOS;

- enabled research communities and data producing agencies to take advantage of and contribute to ANDS utilities. Initial targets include ARC, CSIRO, BOM, GA, ABS, ASSDA;
- enabled universities to take advantage of ANDS utilities and assisted them with local implementations of data utility services. Initial targets include at least the University of Melbourne, the University of Newcastle, the Australian National University, and Monash University;
- made it possible to discover some data collections and items across multiple research domains;
- provided a persistent identifier service for use by repositories and research communities as part of establishing data federation utilities in their own domains;
- established a service to enable researchers to “Publish My Data”;
- developed a roadmap for the progressive delivery of the improvement of existing and development of new utility services, driven by the immediate needs of key stakeholder groups;
- pilot two new services that enable research communities to improve the quality of their data holdings; and
- agreed protocols and SLAs for how current ANDS registries and discovery services will interface with relevant international equivalents.

4.2.6 End of 2010/11 enduring changes

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

- an Australian Research Data Commons framework will exist into which anyone in the ANDS community can contribute the existence and nature of their data, provided that data is in an Australian Research Data Commons integrated store;
- utility services will be available as per the agreed roadmap;
- a considerable number of key repositories of nationally significant information are visible and accessible to the whole national and international innovation community;
- the ANDS Technical Forum exists, has become self-perpetuating, and is the natural broker for communications and consensus on service integration in a data commons; and
- the ANDS Technical Forum will have agreed a technical framework for data federating utilities to facilitate ‘plug and play’ integration of new utilities and services from anywhere in the data commons.

4.3 Seeding the Commons

4.3.1 Program Aims

To improve the fabric for data management and the amount of content (in strategically chosen areas) in the data commons.

To improve the state of data capture and management across the research sector in a highly targeted way (because of resource constraints).

4.3.2 Program Overview

This program will start to seed the Australian Research Data Commons by seeking to make more content available through it. It will do this by initially working in a way that responds to opportunities and needs as they arise, and then moving on to a more targeted assistance model. ANDS is not funded to store data, but to enable data that is stored to be made visible through the Australian Research Data Commons. It will do this by assisting partners who create the data to store, persistently identify, describe and register it for discovery.

In addition to the work described below, some of the NeAT-funded projects will also bring data into the Australian Research Data Commons in addition to their work on new service development described under the Utilities program. Specific examples for 2009/10 are *The Australian Node of the Human Variome Project*, *Transforming performing arts eResearch*, *Establishment of the National Criminal Justice Research Data Network*, and *Ontology driven Data Management for eResearch*.

The overall engagement model for Seeding the Commons is based around the ANDS teams that will be placed in the field to engage closely with key stakeholders across all ANDS programs. How this works out in practice will differ across the themes outlined below.

There is a range of ways in which ANDS will identify groups with which to work. These approaches are not mutually exclusive. One way will be to engage with NeAT round 2 bids that were not successful. In many cases, the bids had a data component, and the fact that – someone was prepared to put the effort into a bid suggests at least a level of commitment. Another way is to partner with key NCRIS capabilities, particularly those that are data-intensive and were funded after the creation of Pfc. These include TERN, ABIN and IMOS. The process for selection will need to be clearly identified (and agreed by the ANDS steering committee), as well as demonstrably open.

4.3.3 Program Themes

Responding to opportunities and needs

This theme will concentrate on the recruitment of existing content into repositories, identifying existing repositories of useful content, and making all that content discoverable through the Australian Research Data Commons.

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. The technologies used will be a mix of ANDS services, local systems, and ARCS-hosted offerings. If demand for this assistance exceeds available capacity, ANDS will develop a transparent process for allocation of ANDS resources in consultation with DIISR.

Where repositories (or federations) already exist, this theme will assist with their integration into the Australian Research Data Commons. 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 Building Capabilities program).

Targeted assistance

The funding available for ANDS is insufficient to provide a data management solution across the entire research-producing sector. As the marginal cost of working with people is high, this program will need to be highly targeted. The aim of the program is to ensure that the data and metadata generated within the program targets is captured, stored, persistently identified and made accessible (initially at collection level) through the Australian Research Data Commons.

ANDS will target automated/semi-automated data and metadata capture to simultaneously improve the quality of what is captured and increase the quantity (by reducing the barriers to capture). This capture and storage will be undertaken in a way that is preservation-aware. Finally, ANDS will aim to share lessons learned and examples of best practice across the sector.

Note that any engagement will need to be undertaken with an awareness of the tensions between international disciplinary practice and national or institutional mandates.

Content systems enhancement

This theme will support new data management activities in institutions through the provision of ARCS-hosted storage offerings that are integrated with the available ANDS utilities. It will also work with 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.

4.3.4 Proposed Activities for 2009/10

Responding to opportunities and needs

- work with NeAT round 2 bids to see what ANDS can do to help, focussing in particular on those who have been advised that ANDS is especially interested in them;
- work with an initial group of universities (at least Newcastle, UniSA, ANU, Monash) on data recruitment; and
- identify as much content as possible that is already in repositories (or can easily be loaded) and make it discoverable.

Targeted assistance

- partner with ABIN, IMOS, TERN and ALA to help them manage and expose their data;
- fund the University of Melbourne to create a series of ANDS data pages for ISO 2146 entities to demonstrate the value of this approach; and
- undertake detailed planning for work to commence in ANDS 2010/11.

Content systems enhancement:

- convene (together with the Utilities program) an Australian Research Data Commons Technical Forum (ARDCTF);
- map the Australian research data store landscape;
- partner with ARCS to define how ARCS storage and ANDS services will interact, and develop software (or commission its development) to implement this as required;
- develop relationships with equivalent activities overseas to share approaches to data management systems that can inform ANDS; and
- work with DCC, NSF and SURFNET to collaboratively develop tools to help improve the quantity and quality of repository content.

4.3.5 End of 2009/10 outcomes

By the end of 2009/10 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;
- selected the initial targets for the targeted assistance activity and commenced activity; and
- mapped the Australian data repository and data collections landscape.

4.3.6 End of 2010/11 enduring changes

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

- more accessible data is stored in more institutions and is discoverable through the Australian Research Data Commons;
- particular target groups have significantly improved their data management practices;
- the Australian Research Data Commons has much better coverage of a number of identified strategic data resources; and
- 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.

4.4 Building Capabilities

4.4.1 Program Aims

To improve the level of capability for data intensive research (and associated technologies) across Australia by working with research groups and partnering with willing institutions to improve core data competencies.

4.4.2 Program Overview

As cohesive networks of research data are increasingly regarded as an important and enduring part of the collaborative research infrastructure, this program will focus in particular on building the capability of researchers and support staff to contribute to and better exploit national data infrastructure.

The Building Capabilities program 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. It will work with research communities and local e-Research support services to improve particular data-related competencies. It will enhance and add national focus to institutionally based support, materials development, and training initiatives.

ANDS will supplement the activities of this program in the community by establishing a reference group drawn from research communities, data networks, e-Research support services, and data stewards. This group will assist ANDS to identify the key research data competencies required in the community and identify various means to improve them.

This program will provide 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 in integrated engagement activity with staff from all other ANDS programs.

4.4.3 Program Themes

Community Building

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. This engagement, through the Australian Research Data Commons Content Forum (ANDS Content Forum) will be an opportunity to build a community of those who have a stake in building the capability of Australian researchers to contribute to and get benefit from national data infrastructure. ANDS expects to inform its own activity through this forum as well as facilitate some cohesion within the sector.

Capability Building

The Capabilities Program will work in association with other ANDS programs and the ANDS content forum to identify and produce documentation of practical use to those engaged in data creation, analysis, curation and preservation. It will also work with these groups to document community-accepted standards for data retention, sharing, curation, analysis and access services.

The Program will work to improve capabilities through one-to-one interaction with researchers and their teams in the field. It will also provide opportunities for capability development through workshops, seminars and other forums.

The ANDS Content Forum will assist ANDS to develop a program of capability maturity modelling for these services. This activity will lead to a voluntary repository rating system as well as a self-audit and certification scheme that will assist institutions and research funders to define “appropriate” repositories for mandated data deposition.

4.4.4 Proposed Activities for 2009-10

Community Building

- use the ANDS Content Forum to discuss, inform and create guidelines for translating data management policy into practice; and
- engage the community on capacity and capability constraints to be addressed and the means of addressing them.

Capability Building

- develop targeted documentation to support the improvement of data management, analysis, curation and preservation;
- provide access to documentation through various media and through direct contact to support widespread access and use;
- provide opportunity for capability development through workshops, seminars and other forums;
- partner with two NCRIS areas – expected to be ABIN and IMOS - to address specific competencies required to take advantage of NCRIS data facilities (materials development, workshops, consultancies);
- work with practitioners to identify best practices in data management processes, data repository management, preservation planning, data management practices, roles and responsibilities and skill sets for the research community; and

- develop relationships with equivalent activities overseas to share approaches to capability building that can inform ANDS.

4.4.5 End of 2009/10 outcomes

By the end of 2009/10 ANDS will have:

- developed systematic documentation to support the improvement of lifecycle data management practices across a range of research communities and institutions, tested at Monash University, Australian National University, CSIRO, University of Newcastle, University of Melbourne, University of Sydney, and the Queensland University of Technology;
- established ANDS staff, in distributed locations around Australia, able to target core set of data related competencies in their allocated communities;
- successfully partnered with two NCRIS capabilities to address their data management capability constraints; and
- discussed, informed and created guidelines for translating data management policy into practice through the ANDS Content Forum.

4.4.6 End of 2010/11 enduring changes

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

- selected NCRIS capabilities and have addressed data-related capability constraints and improved key data competencies;
- 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 is an initial group of rated and certified data services;
- there is a community accepted audit and certification framework for good quality services and staff;
- a community of practice with a sense of common national purpose has emerged amongst key stakeholders of the Australian Research Data Commons; and
- a body of community accepted knowledge and best practice is in current usage.

5 Program Engagement Strategies

ANDS succeeds from strong partnerships that enable the changes we seek. Generally partners will have needs that require a response from a number of ANDS programs, and partners should never need to navigate ANDS internal structures. Moreover ANDS 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 is developing 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.

ANDS engagements will be focused on partner needs, which will typically cross programs. An ANDS engagement with an external client will therefore be an integrated combination of resources from any or all of the programs. For example, an ANDS team engaging with a large research organisation will need to include expert staff from Frameworks, Seeding the Commons, Capabilities and Utilities to meet this partner's needs in areas such as policy, planning, data management, expertise, storage, stewardship, informatics, etc..

ANDS engagements will also be quite varied in nature. We are engaging with partners to directly change research data practise at universities, NCRIS Capabilities, Publicly 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. AeRIC, ARCS and NCI are important partners in delivering the vision of the Platforms for Collaboration, and ANDS will continue to work particularly closely with ARCS, especially in regards to its data fabric and associated services. Finally and most importantly ANDS is engaged with government through the Department of Innovation.

The forms of engagement will be equally varied – in some cases an ANDS staff member will 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 ABIN and IMOS. In some instances we will have several staff work intensively but for a short period of time with a partner such as is the case 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 it is seeking cultural change, not just technical change, so personal engagement and relationships are important. ANDS expects to appoint staff in most states working on the ANDS relationships based in that state. It is expected that arrangements for these staff will be developed in conjunction with ARCS.

5.1 Program Participants

In addition to the broader strategy of basing staff at MARCS to work on ANDS relationships within the states, each ANDS Program will also be open to the support and involvement of the broader research sector in ways that are appropriate to the focus of the program.

5.1.1 Developing Frameworks

ANDS will contribute strength to the development of data frameworks by resourcing a team within the core participants dedicated to the improvement of data management frameworks. It will also:

- engage additional expertise from where it exists in the sector
- part fund professional development positions in its team to be filled by sector participants on a competitive selection basis
- share the workload of broader and related data framework developments with other interested parties who have appropriate expertise and commitment

5.1.2 Providing Utilities

ANDS will establish the core utilities needed to form the data commons through processes managed by the core participants where:

- technical fora and working groups will be open to the sector
- appropriate providers will be contracted to implement utilities
- sector participants will be sponsored to pilot trial services where their expertise is required and they have an interest
- any discipline utilities that are enhanced and deployed through NeAT Projects will involve all relevant parties needed for success

5.1.3 Seeding the Commons

This program is specifically intended to broaden the involvement in building the data commons. To this end:

- technical fora and working groups will be open to the sector
- the selection process will allow all entities interested in research data management to propose a case for focussed assistance
- ANDS will also sponsor necessary developments for organisations willing and able to immediately contribute data to the commons

5.1.4 Building Capabilities

ANDS will establish a core group to engage in significant outreach to improve organisational approaches to data management. It will also:

- sponsor content fora and working groups open to the sector
- sponsor expertise development nation wide
- part fund professional development positions in the team
- partner with institutional support providers for the development and delivery of support materials

5.2 Targeted Program Engagement Partner Summary

<i>Partner</i>	<i>Programs</i>
IMOS	Frameworks, Seeding, Capabilities
ABIN	Frameworks, Seeding, Capabilities
TERN	Seeding
ALA/IBS	Seeding, Utilities
Monash University	Frameworks, Seeding, Capabilities
ANU	Frameworks, Seeding, Capabilities
CSIRO	Frameworks, Utilities, Seeding
University of Newcastle	Frameworks, Seeding, Capabilities
University of Melbourne	Frameworks, Seeding, Capabilities
University of Sydney	Frameworks, Seeding, Capabilities
University of South Australia	Frameworks, Seeding
Queensland University of Technology	Frameworks, Seeding, Capabilities
Australian Research Council	Utilities
Geosciences Australia	Utilities
Australian Bureau of Statistics	Utilities
Bureau of Meteorology	Utilities
Australian Social Sciences Data Archive	Utilities

5.3 Engagement with Government about data

Many areas of research are heavily dependent on government data – from cadastral data to economic data to government organised surveys. Many issues of data custody are shared challenges – preservation, access, and description are three important examples. For instance, it would be unfortunate if research data managers used a completely different approach to persistent identification to government data managers, as the drivers and benefits are similar.

For all of these reasons there is a need for very close relationship between ANDS and the government agencies that are a major custodian of data or that are influential in data policy. Strong relationships must be maintained with the ABS, NLA, NAA, GA and AGIMO, for example.

5.4 Other forms of engagement

Beyond the engagement at the level of program activities, ANDS will need to undertake an ongoing program of informal engagement with key stakeholders. These will include:

- Senior Staff in the Department of Innovation
- Executive Director, AeRIC
- Executive Director, ARCS
- Leaders of NCRIS Capabilities
- Chair of the Council of Australian University Librarians (CAUL)
- Chair of the Council of Australian University Directors of IT (CAUDIT)
- The Australian Access Federation (AAF) Pilot

6 Governance

The Governance and Management arrangements for ANDS are described in a separate Collaboration Agreement. These arrangements have been deliberately designed to ensure that the governance is as open as possible, consistent with the acceptance and management of risk by the lead agency. Particular elements that may be of interest are detailed below.

6.1 Steering Committee

The ANDS Steering Committee should comprise 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 will be invited to nominate a non-voting observer.

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

6.2 Management structure and responsibilities

6.2.1 Management structure

ANDS management and operational arrangements will

- provide the basis for a high degree of interconnectedness, cross-participation and collaboration between the ANDS programs
- allow flexibility of planning and activity in a changing environment
- clearly articulate portfolio and position relationships
- encourage all staff employed to carry out ANDS activities to have high levels of engagement with ANDS

ANDS staff in all institutions will be encouraged to work collaboratively with other ANDS staff and will be given opportunities to meet regularly.

6.2.2 Executive Director

The Executive Director will be appointed by Monash following the selection process agreed by the ANDS Establishment Project and DIISR.

The appointment must be approved by the ANDS Steering Committee.

The position reports to the Chair, ANDS Steering Committee, and is located at a Monash campus.

6.2.3 Deputy Directors

Deputy Directors will be appointed by the host institution of the relevant program in consultation with the Executive Director and in accordance with a selection process approved by the Steering Committee. Initially there will be three Deputy Directors: two to be located at Monash, and one to be located at ANU.

Deputy Directors will 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.

Deputy 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 Deputy 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.

6.2.4 Program and support staff

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 Deputy 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 Deputy 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.

6.2.5 Key Risks and Risk Management Strategies

ANDS has a Risk Register. The risk assessment methodology, adapted from the Australian Risk Management Standard AS/NZS 4360:2004, involves identifying the risk 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 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 ANDS among funding agencies and influential groups leading to lack of buy-in

Risk Factors:

- lack of confidence in governance, management, or program delivery
- perceptions of slow engagement with areas of the sector
- change of policy with regard to the policies around publicly funded research data
- the ANDS message and brand are diluted by its many partnerships

Risk Mitigations:

- ensure that the Program reflects the Government's expectations through constant dialog
- 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
- ANDS will clearly articulate its message and brand

Risk 2 - ANDS programs are not managed effectively

Risk Factors:

- lack of effective mechanisms for planning, leadership and management
- ANDS staff see their primary loyalty as being to their employing organisation not ANDS
- collaboration between ANDS programs and across locations is not effective
- insufficient ANDS staff to manage and coordinate work
- state based staff will have mixed allegiances

Risk Mitigations:

- management and planning processes are put in place to ensure the efficient conduct of the programs
- regular meetings of ANDS staff are held to build a team approach
- staffing levels are monitored and adjusted as required
- contracts between state based organisations that host ANDS staff will be developed that ensure that staff are clear about their role

Relationships

Risk 3 - ANDS external stakeholders are not effectively engaged

Risk Factors:

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

Risk Mitigations:

- maximise the effectiveness of connections between ANDS 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 wide consultation about the ANDS Business Plan both before and after the commencement of ANDS
- 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 ANDS should include effective stakeholder engagement

Risk 4 – ANDS service providers do not contribute effectively to ANDS

Risk Factors:

- lack of effective arrangements in place to ensure the required utilities services are provided to an agreed service level
- service providers see themselves as disconnected from ANDS 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
- involvement of service providers in planning sessions

Impact

Risk 5 – Data providers/federators do not make their data available through ANDS

Risk Factors:

- researchers do not wish to share their research data
- researchers do not trust the ANDS 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:

- link research funding to the provision of data via ANDS as it becomes available
- recognise researchers through peer feedback for the deposit of data into ANDS via increased citation – would need to be recorded and measured as a performance measure by ANDS
- provide targeted assistance to data federations to assist with integration into the Australian Research Data Commons

Risk 6 – Re-users of research data do not use ANDS-supplied mechanisms to discover and access it

Risk Factors:

- the various strategies for exposing data in ANDS do not result in the data being easily discoverable
- the 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 ANDS accessible data (see section 2.4)
- work with ARCS and the Access Australia Federation to identify a simple set of standard access control policies
- ensure that it is easy to re-purpose ANDS accessible data

Resourcing

Risk 7 – High quality ANDS Staff are hard to recruit

Risk Factors:

- limited availability of skilled staff to perform roles in ANDS
- limited tenure roles potentially on offer within ANDS are not attractive to candidates
- second rate 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)
- investigate non-traditional sources of potential staff

Risk 8 – Funding for ANDS is inadequate to achieve its objectives

Risk Factors:

- federal budget constraints and/or competition from other programs results in reduced funding
- ANDS fails to manage expectations of what can be achieved within allocated budget
- aspects of ANDS can't be realised below a certain funding threshold

Risk Mitigations:

- ensure that ANDS delivers early wins to generate momentum
- manage expectations through frequent communication via a range of channels
- focus ANDS programs to improve likelihood of successful outcomes

7 Promotion

The Australian Research Data Commons will occur if the management of research institutions support it, the information infrastructure that researchers rely on supports it, and researchers use it. Each relevant group needs targeting for promotion. ANDS will 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:

- National Collaborative Research Infrastructure Strategy Capabilities Leaderships
- Universities Australia
- potential ANDS Service Providers and ANDS Data Providers
- the ANDS Policy, Technical and Content Forums
- eResearch related forums and conferences

ANDS will communicate with researchers, data centre managers, repository managers and interested others through a network 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 be making direct and targeted approaches to appropriate staff members at the following overseas institutions to discuss collaborative activities:

- The Joint Information Systems Committee
- The Digital Curation Centre
- SURFNet/SURFfoundation in the Netherlands
- Relevant New Zealand instrumentalities
- The National Science Federation (NSF)
- NSF-funded DataNet projects

ANDS will also be exhibiting and/or speaking at appropriate conferences.

8 Access and Pricing

8.1 Access and Pricing

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

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.

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.

8.1.1 Developing Frameworks

The results of the developing frameworks program will be made available as public documents, on a no warranty, royalty free basis. One of the creative commons licences is likely to be selected.

ANDS will not provide a general data management consultancy service.

8.1.2 Providing Utilities

ANDS Discovery Services

The generic ANDS Discovery Services will be available to the public through the web.

Lodgement will be restricted to research data sets, and therefore subject to review by ANDS, however the only cost of lodgement will be where organisations need to provide suitable support for queries generated through the services.

Persistent Identifier Service

Minting of identifiers will be initially restricted to authenticated users, preferably through the Australian Access Federation, and provided cost free.

Resolution of identifiers to addresses will be a public service, also cost free.

Collections Registry

For access and pricing, the collections registry is a specialised component of the Discovery Services.

Technical Consultancy

The provision of commons integration services in 2009/10 will be at no cost.

8.1.3 Seeding the Commons

The services arising from the content recruitment aspect of this program will be limited to data sets that meet the general ANDS access and pricing goals described above.

The focussed assistance component of this program will also prefer data-sets that meet those requirements, however the specific needs and traditional arrangements applied by communities will form part of the evaluation criteria and once accepted those arrangements will be supported by ANDS.

In particular, this activity is expected to include support for restricted access datasets, so that the data commons will not be uniformly available to the public or to all researchers.

ANDS does not intend to provide the means to support fee-for-service access.

8.1.4 Building Capabilities

The provision of targeted documentation will be made available as public documents, on a no warranty, royalty free basis.

Any seminars/workshops/conferences will be on a full or partial cost-recovery basis.

8.2 Level of usage

Usage will be measured in terms of the take-up of provided capacity and measures of the level of coverage of potential data sources achieved over time. Estimates of anticipated usage are provided in the KPIs.

9 Financial and Human Resources

In addition to NCRIS funding, significant resource contributions from the ANDS constituency will be essential if success is to be achieved. ANDS will work with a range of partners typically by way of teams drawn from staff from within ANDS and the partner organisation. Little of this will appear as cash transfers, but will typically be via co-investment.

In a typical partner engagement ANDS would place a wholly-funded ANDS staff member within an institution to achieve the aims of that institution and of ANDS. Another form of engagement is where ANDS funds the development of a new service or deliverable through NeAT or direct commissioning. We expect that approximately 50/50 co-investment will occur for NeAT activities, where ANDS will co-fund its NeAT investment with ARCS. These factors suggest that ANDS should be able to report an aggregate co-ordinated and focussed effort more than double that provided by the direct funding.

The expected allocation at this point of NCRIS funds across the three years is shown in Table 1 below (all numbers in \$M).

	2007-2008	2008-2009	2009-2010	2010-2011	Totals
ANDS Central	0.61	0.48	0.50	0.59	2.18
Frameworks	0	0.30	0.32	0.40	1.02
Utilities	0	1.54	3.45	3.88	8.87
Capabilities	0	0.48	0.52	0.59	1.59
Commons	0	1.69	4.03	4.62	10.34
Budgeted/actual expenditure	0.61	4.49	8.82	10.08	24.00
Revenue	3	5	6	10.00	24
Carryforward	2.39	2.9	.09	0.0	

Table 1: ANDS expected budget allocation

NOTE: The amount required to fund the NeAT Projects approved for commencement in 2008-9 as well as anticipated extra projects in later years was originally broken out separately in *Towards the Australian Data Commons*. In the above table it has been split 50/50 between Seeding the Commons and Providing Utilities. This is because the outcomes of those projects will mostly either lead to new utilities services, or will be closely aligned with activities in Seeding the Commons. For future NeAT rounds, ANDS will be much more explicit about the kinds of NeAT developments it wishes to encourage, in line with strategically identified gaps in service offerings and discipline coverage. The anticipated NeAT projects and the total amount expected to be spent by ARCS and ANDS in 2009-10 is shown in Table 2.

Budget	2009-10
Aus-e-Lit - Collaborative Integration and Annotation Services for Australian Literature Communities	\$400,000
DATA-MINX - A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities	\$600,000
DIAS-B - Data Integration and Annotation Services in Biodiversity	\$400,000
MACDDAP - Marine and Climate Data Discovery and Access Project	\$400,000
SISS - Spatial Information Services Stack	\$400,000
ASeSS-VO - ASSDA Services for e-Social Science	\$400,000

BioNEAT – Bioinformatics Services	\$250,000
The Australian Node of the Human Variome Project	\$250,000
Australian Parameter Estimation Service	\$250,000
Emergency Animal Disease Bioresponse Collaboration Platform	\$250,000
Transforming performing arts eResearch	\$250,000
Large Scale Temporal Spatial Ecosystem Digital Information Service	\$250,000
Establishment of the National Criminal Justice Research Data Network	\$250,000
Remote Computed Tomography reconstruction and visualization service at the Australian Synchrotron	\$250,000
Ontology Driven Data Management for eResearch	\$250,000
Possible Datamining proposal	\$250,000
Possible Humanities and Social Science collaboration environment proposal	\$250,000
TOTAL	\$5,350,000

Table 2: NeAT Funding by Project

Based on the cashflow for 2009/10, and assuming current interest rates, the estimated interest income over 2009/10 should be approximately \$70,000.

The estimated expenditure of NCRIS funds by organisation in 2009/10 is shown in Table 3.

Program Staff Expenditure at Monash	\$2,005,000
ANDS Central Expenditure at Monash	\$590,000
Program Staff Expenditure at ANU	\$1,510,000
ANDS Central Expenditure at ANU	\$120,000
Total internal expenditure	\$4,225,000
NeAT-funded activity	\$2,675,000
Other external expenditure	\$1,920,000
Total external expenditure	\$4,595,000
Total expenditure	\$8,820,000

Table 3: Expenditure of NCRIS funds by organisation

The remainder of the budget will be spent outside Monash and ANU on specialist subcontracting, outsourced staff and NCRIS Projects.

The expenditure of NCRIS funds by type in 2009/10 is shown in Table 4.

Category	Amount	Notes
Staffing	\$7,408,000	The bulk of the budget will be spent on direct staff costs, both inside and outside the ANDS institutions, as well as contractors
Administration	\$300,000	Required to pay for provision of services such as HR and Finance, as well as leased computers, communications charges, office space etc.
Physical Infrastructure	\$0	ANDS has currently no commitments to physical infrastructure in the conventional NCRIS sense.
Related costs	\$1,112,000	Event coordination for the various fora, as well as travel and miscellaneous expenses for the Executive Director, Deputy Directors and program staff, and external service hosting

Table 4: Expenditure of NCRIS funds by type

The types of staffing positions that will be funded under the first year of ANDS include the following:

- events/community coordinators
- policy specialists
- technical analysts
- business analysts
- project managers
- software developers
- system integrators
- metadata experts
- instructional designers
- managers

10 Milestones

10.1 Milestones for 2009-10

The main milestones for ANDS in 2009-10 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 are:

- discussions with key public sector agencies on clarifying access to data continued and some data brought into the Commons (through collaboration between the Frameworks and Seeding the Commons programs)
- developed guidelines on how to comply with the data management requirements in the Australian Code for the Responsible Conduct of Research through an Australian Research Data Commons Policy Forum, and trialled these within an institution
- further Australian Research Data Commons Policy Forums conducted (ANDS Policy Forums) to discuss, inform and create model policy in specific areas such as data management processes in institutions and data-sharing policies
- trialled a model institutional data management plan within an institution
- increased the number of institutions with research data management plans (through collaboration between the Frameworks and the Building Capabilities programs)
- provided written advice on data management to the Research Training Scheme (through collaboration between the Frameworks and the Building Capabilities programs)
- provided governance, policy and other support for the creation of a discipline based data sharing initiative
- ANDS Discovery Services make it possible to discover data collections and items across multiple research domains (although access to the collections and items may need to happen through native interfaces)
- persistent identifier service available for use by repositories and research communities as part of establishing data federation utilities in their own domains
- uptake of ANDS services by NCRIS areas, research institutions, and research communities
- protocols and SLAs for how current ANDS registries and discovery services will interface with relevant international equivalents agreed
- roadmap developed in response to research community needs for the progressive delivery of utility services
- ANDS Publish My Data service established (pilot, then production)
- ANDS data quality services established (pilot)
- Australian Research Data Commons Technical Forum (ANDS Technical Forum) continues to provide advice to the Seeding the Commons and Utilities programs on technical directions

- 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
- selected the initial targets for the targeted assistance activity and commenced activity
- Australian Research Data Commons Content Forum (ANDS Content Forum) has discussed, informed and created guidelines for translating data management policy into practice
- first version completed of a comprehensive set of materials targeting core data-related competencies
- mapped the Australian data repository and data collections landscape
- partnership established with at least four institutions to better capture, manage and share their research data
- assisted IMOS and ABIN to achieve their data objectives through concerted and coordinated efforts from all ANDS programs
- enabled researchers to access some significant government data through the Australian Research Data Commons

10.2 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.

10.2.1 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 our 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 our 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 it 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 our 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

The Institutions we will be monitoring for measurement will focus on the 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 there needs as well – this includes funders, assessors, institutional representatives, such as DVC-Rs, eResearch Directors, Information providers such as libraries, IT providers such as University ITS Departments, partner service providers, such as ARCS and NCI, as well as umbrella organisations such as disciplinary bodies such as the Academies, international research bodies, etc.

The intention of these measures is to provide a baseline in the first year – to measure what is current practise, and enable ANDS to measure progress in subsequent year. Each year as part of the business planning process, there would be a determination of what would represent satisfactory progress for that year.

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

10.2.2 Key performance Indicators for 2009-10

1. The number and coverage of data repositories providing metadata feeds to the national registry compared to the number of data repositories : we intend to build at least 20

automatic plus 40 manual metadata feeds. This is intended to cover at least 7 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: We intend to engage with all Universities, PFRO's, and major Government data providers this year, and to engage at least 20 research groups directly
3. The number of institutions with research data management policies and practices consistent with ANDS recommendations: 3
4. The number of times a search is initiated with an ANDS discovery service: 0 – we expect that in this year we will concentrate on supporting Google discovery of ANDS data pages, and “see also” services.
5. The number of times an ANDS data page (defined below) is accessed : 36,000 – this is based on a ramp-up as there will be few pages that can be discovered at the start of 2009-10, but we will concentrate on directing traffic that way as the pages increase in number
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 our stakeholders – principally research institutions, government data agencies, government research agencies : we aim to strike at least 7 agreements to make data available
8. The number of research data sets in harvestable repositories: 8000, up from the current figure of 4095
9. The number of research data sets with persistent identifiers : 8000, up from the current figure of 20 – thus all data sets that are harvested will be persistently identified.
10. The number of times a data set is reused and referenced – the ultimate long term measure

11 NeAT Projects

11.1 NeAT Status

11.1.1 NeAT Round 1 Projects underway

During 2008, a number of projects under the National eResearch Architecture Taskforce banner were approved, with commencement in 1 July 2008. The following are those projects:

- ASeSS ASSDA Services for e-Social Science
- AustLit Collaborative Integration and Annotation Services for Australian Literature Communities
- DataMINX A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities
- DIAS-B Data Integration and Annotation Services in Biodiversity
- MACDDAP Marine and Climate Data Discovery and Access Project
- SISS Spatial Information Services Stack

The business plan summaries for these projects from the ARCS and ANDS business plans can be accessed at <http://www.pfc.org.au/bin/view/Main/NeATprojects>, and are appended below.

11.1.2 Round 2

During October 2008, a further open call for ideas for NeAT projects was made. Over November and through to the end of January, NeAT considered those ideas and assigned each of them one of five possible outcomes:

- A contact point with ANDS and ARCS to progress the idea further for project planning
- An invitation to work further with NeAT on distilling a project from multiple ideas
- A contact point to consider how the idea relates to existing NeAT projects
- A contact point with ANDS or ARCS to review how the idea might be addressed by, or influence, their own developments
- An invitation to work with ANDS and ARCS to see how research identified in the idea could be assisted

The ideas leading to project proposals being scoped are listed below:

- Australian Node of the Human Variome Project
- Australian Parameter Estimation Service
- Emergency Animal Disease Bioresponse Collaboration Platform
- eResearch service around AusStage
- Large Scale Temporal Spatial Ecosystem Digital Information Service
- National Criminal Justice Research Data Network
- Ontology Driven Data Management for eResearch

- Remote Computed Tomography Reconstruction and Visualization Service at the Australian Synchrotron

The original submissions around these ideas are also appended to this business plan. These were explicitly required to be short two-page documents. They are being worked up into more detailed proposals, but these are not yet complete.

In addition:

- A biotechnology proposal from the first round of NeAT project development is still progressing
- The need for a data mining service will be explored further
- The services that might support collaboration in the humanities will be explored further

11.2 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.

NeAT Project Cost Division: ARCS and ANDS have discussed how the funding for NeAT projects should best be divided between ARCS and ANDS in the coming business year. The following have been agreed as principles to guide the funding allocation:

- All the proposed NeAT projects have both ANDS (data management) and ARCS (collaboration tools and services) related components; as such, both entities are interested in their success and care about their governance;
- All the proposed projects will have ANDS and ARCS representation on their steering committees;
- The precise nature of the projects (and the relative emphasis of ARCS or ANDS concerns within the projects) will not become clear until after the projects have commenced;
- The money in the NCRIS budgets for ANDS and ARCS should be seen as Pfc money to be used for the greater good of NCRIS;
- ARCS and ANDS have developed a close and collaborative working relationship;
- ANDS and ARCS have agreed that the NeAT projects be funded 50% by ANDS and 50% by ARCS. This funding split will be reviewed annually as part of the normal business planning cycles for both ARCS and ANDS.

11.3 Spatial Information Services Stack (SISS)

11.3.1 Service Description

Description of a research community and the eResearch service need

The complexity of data integration is rapidly increasing (more data sources and more combinations of interest) and traditional data integration methods have become untenable as the difficulty exceeds the available human resources. There is increasing need for researchers to access data and services based on open standards for interoperability.

Spatial data infrastructure is being developed by NCRIS capability areas such as AuScope, IMOS and Atlas of Living Australia, as well as many other spatial data providers including the CSIRO Water Resources Observation Network, SEE Grid and the Western Australian Shared Land Information Platform (SLIP). This infrastructure should be interoperable, with common tools and services conforming to Open Geospatial Consortium (OGC) and ISO

standards, in order to establish a spatial information data commons that will be utilized by a variety of research communities.

Description of the proposed service solution and how it meets that need

The SISS project will develop some of the software components, services and functional capabilities needed to realise a spatial information data commons within Australia, conforming to Open Geospatial Consortium (OGC) and ISO standards. SISS will include complete applications or services comprising client interface (e.g. web portal or web service interface), middleware and data repository components. While the software and services that are developed will be generic, the project will focus primarily on addressing the needs of the geosciences community, particularly AuScope.

AuScope will use SISS to deploy an OGC Catalog Service providing a web service interface to a registry with multiple registers to support both discovery of data and governance of standards associated with the spatial data commons (e.g. interoperability profiles, controlled vocabularies, data standards), and an associated spatial information discovery portal. These will be hosted by ARCS.

AuScope will provide SISS-based access to a number of spatial data repositories including GPS network station and observation logs, Seismic and Hyperspectral transects, Geological models, maps, and the Virtual drill core library. The members of AuScope holding the relevant data will sustain these services into the future.

The SISS project will build a skill base able to work with holders of spatial data to deploy and operate relevant data servers and OGC compliant services and a consequent increase in the number and variety of spatial data sets made available through common access mechanisms.

11.3.2 Benefits and proposed measures

Benefits to the user community and associated measures

The SISS should reduce the per unit cost of data publishing and access, particularly for research requiring data from a diversity of domains. Many research activities require access to data held by government agencies and this stack will facilitate improved access to spatial data of interest. Research addressing larger problems (systems, cross-disciplinary research) with more meaningful questions and answers should be possible.

Monitoring of the use of SISS-enabled portals and data access services will measure the volume and variety of data that is accessible through standardised mechanisms, and increases in data use, as well as access patterns.

Benefits to ANDS or ARCS (or other provider) and associated measures

ANDS will improve its understanding of methods for accessing loosely coupled spatial data.

There will be an increased use of spatial data hosted by government agencies (e.g. geological surveys), the MARCS (e.g. CSIRO), and other spatial data providers.

Expected flow-on benefits to others

Use of OGC standards and standard software components for serving spatial data will progress the creation of a national spatial data commons. Improved data search and discovery services, and access to data using widely-used OGC standards, will enable the use of data by a range of user communities.

Communities outside AuScope and geosciences that are expected to benefit from the outcomes of the project are very broad, but include IMOS, Atlas of Living Australia,

Terrestrial Ecosystem Research Network, Biosecurity, water resources (CSIRO WRON, AWRIS, e-Water CRC), Australian Bureau of Statistics, social sciences.

11.3.3 Resources and commitments

Resources provided by the user community

AuScope will provide 5 EFT and CSIRO e-SIM will provide 1 EFT of effort working on SISS or SISS-related development effort. An estimated 3 EFT of this effort would be working directly towards the goals of this project.

Resources provided by ANDS or ARCS

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers through IVEC. However this will be decided by the Project Committee and specified in the Project Plan.

Some ARCS Operations effort will be utilized to deploy SISS at MARCS hosting spatial data sets. A rough estimate of this effort is a total of 0.5 EFT p.a. across all the MARCS.

Resources provided by others

Some effort in deploying SISS by a variety of organizations hosting spatial data, including government agencies such as state Geological Surveys, and MARCS such as CSIRO.

Total project resources and commitments are summarized in the following table.

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (AusScope, CSIRO)		3		3		3		9
ANDS	200K		200K		100K		1.0M	1.5
ARCS	200K	0.5	200K	0.5	100K	0.5		

11.3.4 Governance

Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee

- ANDS Executive Director, or nominee
- CEO Auscope, Scott McTaggart (Chair)
- General Manager, Office of Spatial Data Management, Ben Searle
- Geological Survey of Victoria, Alan Willocks
- Geoscience Australia representative (to be announced)
- CSIRO Land and Water Chief, Neil McKenzie, or nominee

The Project Manager is Dr. Robert Woodcock

11.3.5 Project Summary

Deliverables / Milestones

July 2008 - Dec 2008

- ARCS Hosting OGC Catalog Services and Discovery Portal
- Auscope hosting OGC Catalog Service and Discovery Portal
- Auscope GPS data WFS service deployed
- Geological Survey of Victoria GeoSciML testbed collaboration deployed

Jan 2009 - Dec 2009

- ARCS SISS support service fully established
- CSIRO Minerals Down Under Flagship deploys laterite geochemical data service (Western Yilgarn)
- Auscope NVCL WFS service deployed
- CSIRO Minerals Down Under Flagship deploys airborne hyperspectral data service
- CSIRO Minerals Down Under Flagship deploys thermodynamic data service
- CSIRO Minerals Down Under Flagship deploys Northern Yilgarn hydro-geochemistry

Jan 2010 - Dec 2010

- Auscope Virtual Rock Laboratory and Tsunami workflows utilise registry and information service infrastructure for service discovery and data management support
- Auscope deploys WCS for geophysics imagery with large data set support

Jan 2011 - June 2011

- Auscope Earth Model and portal service infrastructure fully established and using the service stack
- Broader adoption well underway

Overall risk assessment

The major risks and their mitigation are:

- Project complexity. Project characteristics are similar to a previous successful project. The proposed project leader has a track record of successfully managing projects of this size.
- Stakeholders cannot agree on a common set of requirements. Communications involve stakeholders in a group situation (the Project Committee and Reference Group) as well as one-on-one. All stakeholders are demonstrating high levels of commitment to the project.

- Unable to secure appropriate and skilled staff for the required work. Can use secondments from partner organisations, and contractors to augment the development of the services.
- Technology development is highly challenging. There is a parallel approach in the project plan to mitigate greater than normal development risk. There is a contingency plan to recover significant value should the project not fully succeed. All stakeholders are fully aware of and accept the chances for project success.

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.4 Marine and Climate Data Discovery and Access Project (MACDDAP)

11.4.1 Service Description

Description of a research community and the eResearch service need

Researchers across marine and climate communities need to better use their combined data resources, as well as instruments to observe physical and biological properties around Australia. Basic services are being established through IMOS. This project supports providers of marine and climate data sets, by creating efficient services built on international standards and software, to more easily manage, translate, and control these distributed digital repositories for the benefit of Australian researchers.

Description of the proposed service solution and how it meets that need

The underlying services that are to be provided by this project are :

Access Services: to enhance availability of marine and climate data in a wider range of standard protocols, including Open Geospatial Consortium (OGC) standards, integrated into OpeNDAP, the standard protocol used for accessing many of these data sets.

Discovery Services: a Metadata Entry Search Tool, an OpeNDAP digital library metadata harvester, a Catalogue Exchange Service and an Aggregation Service will ensure that the distributed data sets are increasingly discoverable and conformable with standard vocabularies.

Support Services: Workflow tools that process existing data sets in order to generate standard metadata and enable the data to be accessed via the OGC services.

The services will be delivered through the MARCS, particularly TPAC and CSIRO, but also the Bureau of Meteorology and others.

11.4.2 Benefits and proposed measures

Benefits to the user community and associated measures

Users will be able to more easily discover marine and climate data, and to access it using OGC standard web service interfaces. Development of automated workflows will reduce the cost of providing access to data and metadata using standard formats and interfaces.

Monitoring of portal use will measure increases in data use, and increases in the use of OGC standard interfaces, as well as access patterns. The impact of the services will be formally measured before and after automated services are created.

Benefits to ANDS or ARCS (or other provider) and associated measures

ANDS will be able to understand and measure the extent to which marine and climate researchers access IMOS. In particular, much of this data is geo-referenced, ANDS will improve its understanding of methods for accessing loosely coupled spatial data.

There will be an increased use of data hosted by the MARCS and accessed through marine and climate data portals hosted and maintained by the MARCS.

Expected flow-on benefits to others

Use of OGC standards for serving spatial data will progress the creation of a national spatial data commons. Improved data search and discovery services, and increased access to data using widely-used OGC standards and OPeNDAP protocol, will enable the use of data by a wider range of user communities.

Many important international data providers utilize the OPeNDAP and GeoNetwork software that will be enhanced by the work on this project, and hence will also benefit from the outcomes of this project.

11.4.3 Resources and commitments

Resources provided by the user community

IMOS will provide \$10K per annum cash and 3.2 EFT (2.5 from eMii, 0.7 from remote sensing stream).

Resources provided by ANDS or ARCS

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

Approx 1.5 EFT of ARCS Operations effort will assist AAF support, monitoring support for developed services, and deployment and hardening of services at the MARCS.

Resources provided by others

1.25 EFT from Bureau of Meteorology and CSIRO, 1.2 EFT from TPAC.

Total project resources and commitments are summarized in the following table.

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (IMOS)	10K	3.2	10K	3.2	10K	3.2	30K	16.95
Other (TPAC, BoM, CSIRO)		2.45		2.45		2.45		
ANDS	200K		200K		100K		1M	4.5
ARCS	200K	1.5	200K	1.5	100K	1.5		

11.4.4 Governance

Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- IMOS Executive Director, Professor Gary Meyers (Chair)
- eMii Director, Professor Roger Proctor (or Ms Kate Roberts)
- ARC NESS Convenor (Prof. Andy Pitman)
- Prof. Nathan Bindoff

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

11.4.5 Project Summary

Deliverables / Milestones

Major deliverables for the project are:

- Discovery Services
 - MEST enhancements (GEONETWORK, OPeNDAP)
 - TPAC Digital Library Portal enhancements (data registration, geospatial coordinate aware, gridftp, WCS and WMS service aware, simple visualisation)
 - Enhanced OPeNDAP harvester (geospatially aware, geospatial coordinate aware, gridftp, WCS and WMS service aware, simple visualisation)
 - Aggregation Service for remote sensing
- Access Services
 - OPeNDAP data delivered as WMS, WCS and trialled with AAF
 - OPeNDAP enhancements (authentication, administration, remote management, and data handler enhancements, server-side functions)
- Support Services
 - Translation Service (web based workflow for OPeNDAP to ISO19115 standards)
- Deployment of the above services at the providers nominated in 1.2.

The enhancements of the software, community profiles, translation tools and related workflows in the proposed services will be delivered back to the international community, through their respective mechanisms or provided on Sourceforge where appropriate

Milestones in the first year of the project are:

- Development of OGC WMS services for OPeNDAP THREDDS server.
- Deployment of updated OPeNDAP servers at sites in Australia
- Prototype Aggregation Service
- Data conversion tools and workflows
- Improvements to TPAC Data Portal
 - Allow searches on spatial information
 - Improve search performance
 - Provide data visualization in portal
 - Services for registering data sets and associated metadata with the portal

Further details of deliverables and milestones are described in the MACDDAP Project Plan.

Overall risk assessment

The major risks and their mitigation are:

- Unable to secure appropriate and skilled staff for the required work. Mitigation is the use of secondments from partner organisations, and contractors to augment the development of the services.
- Projects that depend on each other and cause time delays. This has been mitigated by minimising co-dependencies between projects.
- Uptake of the new services (e.g. translation service for pre-existing data sets so that they can be advertised in the eMii marine catalogue.). The mitigation strategy is built in the plan by allowing funds for the active interaction with staff that hold data and to develop specialised plug-ins for the data providers' particular environment.

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.5 Data Integration and Annotation Services in Biodiversity (DIAS-B)

11.5.1 Service Description

Description of a research community and the eResearch service need

The Atlas of Living Australia (ALA) needs to support integration of a wide range of different types of biodiversity data – taxonomic data (e.g. taxon names and synonyms), specimen and observation data, species descriptions and associated images, diagnostic keys, genomic data, etc – from many different data providers. The user community for the ALA is very broad, encompassing taxonomists, botanists, zoologists, environmental scientists, land-use and conservation planners, and biosecurity officers. The Australian Centre for Plant Functional Genomics will be a specific user of plant phenomic data mediated through the project.

In order to provide discovery and interoperability across many and varied biodiversity data sets, the ALA requires needs best practices for metadata management, including adoption of relevant vocabularies and ontologies, and the ability to map between different metadata models. A Metadata Repository is required to enable metadata registration and harvesting for all available digital resources of biodiversity information.

The quality and consistency of the ALA data is crucial for its use. There is a need for an authenticated annotation service that will allow users or automated data analysis tools to provide information to users and feedback to data providers by annotating data records and resource metadata with comments on data quality and suggested corrections.

Description of the proposed service solution and how it meets that need

Data quality services:

- Annotation service allowing human and machine users to store and retrieve annotations relating to any data record within the ALA to record possible errors.
- Reporting service that alerts data provider/owners of possible quality issue.

Data integration services:

- Catalogue of mandated and supported data standards, vocabularies, ontologies for use within the ALA.
- Metadata repository and metadata registration software for registration of all Australian biological data resources and for relating data sets to supported vocabularies and ontologies.
- Search interfaces (including web service interfaces) to search the metadata repository using terms from the supported ontologies.

Together, these services will enable data providers and researchers to actively participate in the creation and use of the Atlas of Living Australia.

These services will initially be hosted by ALA, however if these services can be made more generic they may be hosted by ARCS on behalf of ANDS.

11.5.2 Benefits and proposed measures

Benefits to the user community and associated measures

Easier and faster integration of new data sources into ALA, measured by:

- Level of provision of metadata to metadata repository
- Number of data sets accessible via ALA
- Number of records accessible through ALA.
- Usage of ALA

Improved data discovery and federated search, measured by:

- Direct use of metadata services from the repository by other networks and repositories
- Percentage of data resources with metadata entries including references to ontology terms
- Extent of discovery through ontologies

Improvement of data quality via use of annotation services, measured by:

- Number and range of annotations in annotation database
- Number of responses from data providers
- Direct use of services (UI and web services) for providing annotations (other than through the ALA portal UI and ALA data validation tools services)
- Direct use of services (UI and web services) for accessing annotations
- Number of data records for which annotations have led to corrections in source data

Improved level of user experience, measured by:

- Independent reviews to be contracted at the end of 2008-2009 and at the end of 2010-2011 to document the experience of key target user groups and to compare the state of ALA infrastructure with other national biodiversity information platforms.
- An online survey tool to allow users to document their experience in using the ALA infrastructure. This survey tool will be continuously available as a data capture method. This survey will explicitly determine success in using data quality annotation.
- Analysis of web logs to determine whether users are guided to relevant information.

The Project Plan and the ALA Business Plan will specify some quantitative goals for these metrics for each year of the project.

Benefits to ANDS or ARCS (or other provider) and associated measures

There are two key benefits being sought. Firstly, an ability to integrate several different data sets with different schemata or ontologies, enabling researchers to find things despite having different knowledge lenses – this will be measured using the surveys described above. Secondly, we wish to understand how well an annotation service supports improvement of data quality – it will be measured using the survey described above.

Expected flow-on benefits to others

As articulated above, this affects all disciplines where commentary on other work is important. More specifically, the services developed within this project would also be of benefit to:

- NCRIS 5.12 Marine Sciences and Climate
- Social Sciences, ASSDA, AustLit
- NCRIS 5.3 Microscopy and Microanalysis
- NCRIS 5.8 Bio-security
- NCRIS 5.11 Terrestrial Ecosystem Research Network

11.5.3 Resources and commitments

Resources provided by the user community

The Atlas of Living Australia will have approximately 5 EFTs working on software development related to this project. The Australian Biological Resource Survey (ABRS) and the Australian Museum also have existing developers whose products will be contributing directly to the development of the Atlas.

Resources provided by ANDS or ARCS

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The actual amount for the third year will be dependent on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers at sites with relevant expertise, possibly including CSIRO, SAPAC, ANU and UQ. This will be decided by the Project Committee and specified in the Project Plan.

Resources provided by others

There will be significant related international effort in standards development by the Taxonomic Data Working Group (TDWG) and open source software development effort from members of TDWG, GBIF and EoL (the Encyclopedia of Life). It is expected that some of

these standards and software will be utilized in the NeAT project. This effort is difficult to quantify and not included here.

Total project resources and commitments are summarized in the following table.

	Y1		Y2		Y3		Total	
	Cash	EFT	Cash	EFT	Cash	EFT	Cash	EFT
ALA		5		5		5		15
ANDS	200K		200K		100K	1	1.0M	1
ARCS	200K		200K		100K			

11.5.4 Governance

Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- Director of the Atlas of Living Australia, Donald Hobern, Chair
- Director of e-Research at the University of Queensland, Dr. Jane Hunter
- Federation Fellow, Professor Hugh Possingham

The Project Manager will be Dr. Lynette Woodburn

11.5.5 Project Summary

Deliverables / Milestones

Metadata Repository Activities - 2008/09:

- Review of metadata management and requirements in related international biodiversity informatics projects (particularly GBIF and EOL)
- Review metadata standards and ontologies in use within relevant Australian and international projects, and mappings between them.
- Review available software options for a metadata repository.

Metadata Repository Activities - Years 2-3:

- Contribute to the development of the TDWG core ontology.
- Establish standards for the use of unique identifiers for data resources and data items.

- Develop user interfaces and web services for primary registration of data resources and for configuration of OAI-PMH harvesting.
- Develop user interfaces and web services for search and selection of data resources via ontology terms as well as free-text search.
- Develop alternative output metadata formats (based on review of metadata standards above).
- Investigate how to integrate outputs from the Annotation Service into metadata management.

Annotation Service Activities - 2008/09:

- Investigate requirements for annotation services in other NCRIS capabilities and in ANDS.
- Investigate existing collaborative annotation systems and select the most appropriate solution.
- Investigate how to integrate it with the Metadata Repository and other components in the ALA system.

Annotation Service Activities - Years 2-3:

- Develop an appropriate user interface that may need to be customised for structured annotation of different types of data.
- Test automated annotation of records by error-checking tools.
- Develop interfaces for management of obsolete annotations (e.g. after data record has been corrected for errors) and for threaded annotations (e.g. data provider responses to user comments)
- Test and refine the interface with a variety of users.
- Provide support for AAF authentication.

Overall risk assessment

Risk	Mitigations
Take up is slow	<ul style="list-style-type: none"> • Measure take up • Project leader to be responsible for identifying projects and collaborations to provide and consume metadata and annotations
Poor software	<ul style="list-style-type: none"> • Good people/place – specifically oversight of experts from UQ, SAPAC and ANU
Other approaches are more attractive	<ul style="list-style-type: none"> • Keep watching and be adaptive in project development to make best use of developing standards and practices

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.6 Collaborative Integration and Annotation Services for Australian Literature Communities (Aus-e-Lit) Project

11.6.1 Service Description

Description of a research community and the eResearch service need

AustLit currently provides a Web portal to arguably the most important collection of information for scholarly research into Australian literature and print culture. The Association for the Study of Australian Literature (ASAL) is the peak body for this research community. The current collection primarily supports bibliographic records with some full-text and limited image, audio or video content. The aim is to expand the AustLit collection to include more comprehensive access to full-text works as well as related content (images, recordings, reviews, critiques etc). In parallel, the search interface will also need to be enhanced to enable support for metadata, full-text, empirical and multimedia searching of external sources via a single web portal. Currently a large number of searches through many different interfaces across many data sets is required. As well, there is no electronic support for a core activity of literature researchers – to create and share collaborative commentaries via document markup or annotation tools. Finally, researchers would like to be able to encapsulate related artefacts into e-learning objects that they can share and exchange with other researchers and educators.

Description of the proposed service solution and how it meets that need

The above needs will be met by the creation of several new services:

1. Data integration, search and reporting services
 - OAI-PMH metadata harvesting services that periodically harvest metadata from the specified external repositories and aggregate the metadata with the AustLit metadata repository
 - Federated Search Portal supporting metadata and full-text search across multiple data sets, and search, retrieval and presentation of records, text and images
2. Collaborative annotation services
 - Secure annotation creation, editing and attachment services
 - Annotation presentation, browse and search services
 - Annotation storage and servers (with authenticated access control)
 - Annotation harvesting and aggregation services
3. OAI-ORE compliant compound object authoring, editing and publishing services.

AustLit-specific services will be hosted and operated by the UQ Library, which currently hosts the AustLit server. The more generic services that may be deployed in disciplines other than AustLit will be operated initially by UQ's eResearch group and in the long term by UQ's IT Services and/or ARCS.

11.6.2 Benefits and proposed measures

Benefits to the user community and associated measures

This project will widen the scope of AustLit, providing more comprehensive access to Australian literature resources. The project will also promote and expedite the adoption of eResearch practices by humanities researchers in Australia. Success will be measured by monitoring the number of site visits, searches and references to AustLit, both by researchers, but also by web-links such as from Wikipedia, and scholarly works. The project will also

enable groups of researchers to discover all material relevant to say “the visit of D. H. Lawrence to Australia”, or “the battle of Lawson and Paterson”, to create a collective annotated bibliography that can be explored and added to in a way that the current best practice – a book – cannot. A core research method of most of the humanities, law and the social sciences will be “e-enabled”. Results will be measured by monitoring usage of the services, the creation of new resources, and interviews with researchers on tool usefulness. The AustLit advisory board is derived from active researchers and teachers in the field and they will be involved in the testing and feedback cycle as the project develops.

Benefits to ANDS or ARCS (or other provider) and associated measures

The ability to search seamlessly across both the metadata and the underlying data, particularly text, is an important one for many research areas. Much of the data access is envisaged to be by metadata, yet this is not the case for Google, say – this work will demonstrate how we bridge this boundary. The project will also provide an exemplar of data integration, with a federated search across multiple distributed databases. The only relevant measure here is that researchers use this new search facility and keep doing so. The ability to annotate data and other objects, and groups of objects, is important to many fields, but especially to the humanities. We need to demonstrate that this can be e-enabled – at the moment search is electronic, but not this aspect of the “work” of the humanities researcher. We will measure this principally by use and output. The ability to easily author compound objects in a standardized format that can be shared and re-used will greatly facilitate the use of AustLit and other cross-disciplinary content within eLearning resources.

Expected flow-on benefits to others

As articulated above, this affects all disciplines where commentary on other work is important. More specifically, the services developed within this project would also be of benefit to:

- NCRIS 5.12 Marine Sciences and Climate
- Social Sciences, ASSDA
- Atlas of Living Australia, Terrestrial Ecosystem Research Network (TERN),
- NCRIS 5.3 Microscopy and Microanalysis and NCRIS 5.8 Bio-security

11.6.3 Resources and commitments

Resources provided by the user community

In-kind from AustLit, including Kerry Kilner, and other data providers.

Resources provided by ANDS or ARCS

NeAT funding of \$250K p.a. for 2 years and nominally \$250K for a third year. The actual amount for the 3rd year will be dependent on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers at UQ. This will be decided by the Project Committee and specified in the Project Plan.

Resources provided by others

In-kind from UQ Library for hosting the servers and services, and UQ eResearch group, including Jane Hunter. \$35K p.a. for 3 years from University of Queensland. \$35K p.a. inkind for 3 years from QCIF.

Total project resources and commitments are summarized in the following table.

	Y1		Y2		Y3		Total	
	Cash	EFT	Cash	EFT	Cash	EFT	Cash	EFT
User community resources		0.5		0.5		0.5	210K	3.0
Others (UQ, QCIF)	70K	0.5	70K	0.5	70K	0.5		
ANDS	125K		125K		125K		750k	0
ARCS	125K		125K		125K			

Governance

11.6.4 Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.
- A Reference Group will also be established for consultation.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- ASAL President, Elizabeth McMahon (Chair)
- AustLit Executive Manager, Kerry Kilner
- UQ eResearch, Professor Jane Hunter

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

11.6.5 Project Summary

Deliverables / Milestones

2008/09: Full text search over both metadata and content with a variety of presentation formats.

Integrate 5 key databases, build and support federated search interface: AustLit, SETIS,

Australian Dictionary of Biography Online, Dictionary of Australian Artists Online, and National Library services - National Bibliographic Database (NBD), People Australia, Picture Australia, PANDORA, and Australian Digital Theses DB.

2009/10: Collaborative Annotation creation, editing and publishing services

2009/10: Compound Object authoring and publishing services (encapsulate objects related by theme)

More detailed deliverables and milestones will be specified in the Project Plan.

Overall risk assessment

The key risk of the project is that the tools are not used by Australian literature researchers - this is mitigated by the close engagement with the community and rapid delivery of beta versions for user acceptance.

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.7 ASSDA Services for e-Social-Science (ASeSS)

11.7.1 Service Description

Description of a research community and the eResearch service need

The social science and humanities communities have need of a national e-Research infrastructure that enables researchers to explore their research data using a data commons approach that preserves confidentiality. This project involves setting up the first cohesive Australian e-Research service for the social sciences. The project aims to build a national model that incorporates closely related research domains that forms the foundation for an e-Social Science Virtual Organisation (SSVO).

The Australian Social Science Data Archive (ASSDA) collects and preserves computer-readable data relating to social, political and economic affairs and makes the data available for further analysis. The primary *user* community of ASSDA is Australian academic researchers, doctoral candidates, postgraduates and undergraduates in the social sciences and humanities. However, the services offered by ASSDA are structured in a fashion that also allows researchers and staff from Australian public sector agencies, non-government organisations, the media and, occasionally, the private sector to access 'public use' data sets. Much of the high-value data that researchers would like to access for analysis has ethics and confidentiality requirements. Thus a major challenge presented to ASSDA is the need to balance access with assurance of confidentiality, without requiring labour intensive processes.

Description of the proposed service solution and how it meets that need

The ASeSS project will develop a suite of services to support the SSVO. These will include:

- Data curation software that supports reliable data ingestion, and sets up the appropriate access controls.
- Search tools that support authenticated discovery across archives
- Integration of a suite of analytic tools that enable easy use by social scientists (the tools themselves will be developed outside this project)
- Integration of a suite of visualisation tools, particularly spatially oriented and temporally oriented

The SSVO will need to integrate a wide variety of services, and will need strong authentication and authorization before access and use (in most cases). Providing online access to statistical data analysis tools provides additional value to the data archives and also

supports the ‘unlocking’ of data sets of national significance that cannot otherwise be made accessible due to privacy constraints.

11.7.2 Benefits and proposed measures

Benefits to the user community and associated measures

ASeSS should reduce the cost of data publishing, access and analysis. Social scientists should have access to a suite of tools that are easy to use as standard desktop publishing tools. The key measure will be the extent to which social scientists use this toolset, and the data that the toolset supports. The cost of data acquisition is currently high, so it will be important to ensure that efficient data ingestion and curation occurs.

Monitoring of the use of ASeSS will measure the volume and variety of data that is accessible using web monitoring tools and user surveys.

Benefits to ANDS or ARCS (or other provider) and associated measures

ANDS will improve its understanding of methods for data acquisition and access over a wider variety of data types than is typically held in numerically oriented databases. As well, tools that support the exploration of a combination of textually and numerically oriented data will be developed. Social science data is an important element of the data commons so it is important that it is well represented in the data commons. ARCS is charged with supporting Virtual Organisations, and ASeSS will enable better understanding of VO’s with strong authentication and access control. It should be a good testbed for AAF.

Expected flow-on benefits to others

This tool set should be generalisable and thus be of benefit to many user communities that need analytic and visualisation tools for a wide variety of data types. The VO infrastructure will be of value in any research community where there is a need to collaborate within a restricted environment.

11.7.3 Resources and commitments

Resources provided by the user community

ASSDA will provide 6 EFT.

Resources provided by ANDS or ARCS

NeAT funding of \$400K p.a. for 2 years and nominally \$200K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is anticipated that the NeAT funding for this project would be used primarily to hire software developers through ASSDA. However this will be decided by the Project Committee and specified in the Project Plan.

Some ARCS Operations effort may be utilized to deploy ASeSS at MARCS and other institutions hosting social science data sets.

Resources provided by others

This has not yet been established.

Total project resources and commitments are summarized in the following table.

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total

User (ASSDA) community		6		6		6		18
ANDS	200K		200K		100K		1.0M	
ARCS	200K		200K		100K			

11.7.4 Governance

Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

Project Committee:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- ASSDA Director, Dr. Deborah Mitchell
- Research community representatives, to be determined

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

11.7.5 Project Summary

Deliverables / Milestones

Period	Major milestones
Dec 2008	Demonstration version of GIS data viz web tool Demonstration version of longitudinal data analysis web tool
June 2009	First release of new ASSDA VO web site Demonstration of first text-based VO service deployment Demonstration of cross-archive data search between two major archives Historical Census and Colonial Data Archive (HCCDA) availability
Dec 2009	Time Series & Panel Archive and Qualitative Data Archive available Text based analysis tool demonstration for Qualitative Data Archive Search available on UR data archive
June 2010	Search available for Qualitative Data and HCCDA on VO web site

	<p>Longitudinal data analysis web service available on VO web site</p> <p>Data exchange service for Time Series & Panel Data and Qualitative Data</p> <p>Q/A component of curation software for Qualitative data</p> <p>Production version of web-based UR archive curation service on VO web site</p>
Dec 2010	<p>Time Series& Panel Archive services available on VO web site</p> <p>Q/A component of Indigenous data archive curation service</p> <p>Search available on Electoral Database on VO web site</p>
June 2011	<p>Release production version of web-based Qualitative data archive curation service on VO web site</p> <p>Search available over Indigenous archive</p> <p>Release production version of Indigenous data archive curation service on VO web site</p> <p>Generalised version of GIS service on UR data available on VO web site</p>

Overall risk assessment

- The major risks and their mitigation have not yet been completed.

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports by the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.8 A Data Fabric for Characterisation – Microscopy, Imaging, Neutron and X-ray Facilities (DataMINX)

11.8.1 Service Description

Description of a research community and the eResearch service need

Thousands of researchers use molecular and material characterisation facilities, such as those funded by NCRIS 5.3, including the neutron facilities at ANSTO, the Australian Synchrotron, institutional X-ray diffraction facilities, and the Australian Microscopy and Microanalysis Facility (AMMRF), which is a network of major institutional facilities and linked labs of national significance. These facilities currently do not provide the capability for researchers to easily and reliably transfer experimental data from the facility to remote data storage or a data repository, with automated capture and storage of associated metadata, and the ability to easily provide authenticated sharing of the data to colleagues or to publish the data in a way that it is easily discoverable and accessible by any researcher. Ideally these services should be uniform across the different facilities and leverage the national data fabric provided by ARCS and the data storage and data repositories provided by the MARCS and other institutions. A significant number of researchers who use Australian characterisation facilities also use international facilities, hence interoperability and linkage with international facilities is highly desirable.

Description of the proposed service solution and how it meets that need

The project will develop a “data fabric” of data services and infrastructure to support the above user and facility requirements. The underlying services that will be developed by this project will include: automated capture of data and metadata from instruments and transfer to a data repository; conversion of data and metadata to standard formats; a federated data repository; a searchable metadata catalog; fast data transfer and download; and a web portal to discover and access data, with authorized data access.

The services will build on the ARCS data fabric delivered by the ARCS Data Services team, utilizing data storage provided by the MARCS and other organizations (e.g. VerSI).

It is expected that the services will be based as much as possible on existing software and services, such as those developed by the UK Science and Technology Facilities Council (STFC) for use in the ISIS neutron facility and Diamond synchrotron in the UK, and by the DART and ARCHER projects. These services are based on Storage Resource Broker (SRB) and the ICAT metadata catalog, which implements the STFC Scientific Metadata Model.

The system will initially be targeted at AMMRF, Neutron, Synchrotron and X-ray facilities, but the aim is to extend it to support the Imaging component of NCRIS 5.3 once a working system is deployed and in use.

The detailed specification of the project is still under development and will be specified in the Project Plan.

11.8.2 Benefits and proposed measures

Benefits to the user community and associated measures

An important aim of the project is to move users away from personal, ad-hoc data management and data sharing practices, to the use of use of automated data management services, secure data access and sharing, and institutionally supported long-term archiving of data and associated metadata. The provision of these services will mark a significant change in current community practice, protecting and preserving research outputs, and making it easier for research collaborations to share data and to utilize results from multiple distributed characterisation facilities. This will improve the level of uptake of sound eResearch practice as individual researchers and research groups utilise both the collaborative aspects of the data grid as well as its archival and data re-use/publication aspects.

In addition, characterisation data sources will no longer be viewed in isolation. Instruments will become linked through data grid technologies and services, and this will make it easier for researchers to undertake studies that rely on more than one type of instrument.

The project will monitor the number of experiments and users utilizing the new services, and the amount of data being archived and shared. User community views will be measured by an annual survey.

Benefits to ANDS or ARCS (or other provider) and associated measures

The proposed project provides some large user communities and different use cases for the ARCS data fabric and associated authentication services.

ANDS will be able to understand and measure the extent to which a large distributed research community can make use of services for capturing data and metadata from various scientific instruments and moving it to federated data archives, and the use of authenticated data sharing. The proposed solution can also be used as a solution for institutions that need an integrated data management solution.

Through this project, ANDS will understand how well metadata can be captured with data directly from an instrument that supports both the immediate use, and subsequent uses of characterisation data. As well, ANDS will be informed about metadata transformation for use in an authenticated federated data archives.

Measures will be as described in section 2.1.

Expected flow-on benefits to others

This project will develop a data management system that will support data and metadata capture from instruments, data transfer to federated data repositories, authenticated data access, with an associated metadata catalog that supports a general and extensible scientific metadata standard. It therefore has the potential to be much more broadly applicable.

11.8.3 Resources and commitments

Resources provided by the user community

ANSTO will provide approx 1 EFT of in-kind effort, particularly input into the design and testing of the system, as well as funding for storage and hosting of their data at ac3.

Australian Synchrotron (AS) will provide some in-kind for IT support.

MMSN and Uni of Sydney will provide some in-kind programming effort and input into the design and testing of the system. Expected to be able to provide approximately 1 EFT in the first year, subject to review after that.

AMMRF has funding for 2 new EFTs to support this project, and in-kind contributions from existing staff at the AMMRF facilities for service development and deployment and assistance with managing the AMMRF-focussed effort will be at least another 1.5 EFT.

Resources provided by ANDS or ARCS

NeAT funding of \$600K p.a. for 2 years and nominally \$300K for a third year. The third year amount will depend on the outcome of project reviews and available NeAT funds.

It is expected that the NeAT funds would be used to hire software engineers to assist in customizing existing software and services for particular facilities and user requirements, developing new software and services where required, and working with staff at the experimental facilities to deploy, test and document the services and to assist users in making use of the services. At least one NeAT-funded person would need to be closely associated with each facility – the Australian Synchrotron in VIC, ANSTO and X-ray facilities in NSW, and the AMMRF nodes, which are located in QLD, NSW, ANU, WA and SA.

Significant effort from ARCS Operations will also be provided, through the ARCS Data Services team. This is estimated to be at least 3 EFT. Some effort from the ARCS authorization services project is expected to focus on the requirements of this project.

Resources provided by others

The MARCS have agreed to provide storage capacity at hardware cost.

VeRSI is expected to provide at least 1 EFT.

INTERSECT is expected to be able to provide approximately 1 EFT for the first year, with follow-on to be reviewed in the final quarter of the first year.

The ARCHER project is undertaking work that is closely related to this project. Some ARCHER effort in 2008H2 could be aligned with this project and software developed by

ARCHER could be used. An assessment of ARCHER developments will be made as part of the project.

The project will leverage software developed by the UK Science and Technology Facilities Council (STFC) that is used for the ISIS neutron facility and Diamond synchrotron in the UK. The STFC, ISIS and Diamond e-Science groups are willing to collaborate on the project.

Total project resources and commitments are summarized in the following table.

	Cash Y1	EFT Y1	Cash Y2	EFT Y2	Cash Y3	EFT Y3	Cash Total	EFT Total
User community (ANSTO, AS, AMMRF, MMSN)		5.5		5.5		5.5		22.5
Other (VeRSI, INTERSECT)		2		2		2		
ANDS	300K		300K		150K		1.5M	9
ARCS	300K	3	300K	3	150K	3		

11.8.4 Governance

Governance processes to be applied to the project

- The ARCS/ANDS agreed governance mechanism for NeAT projects, defined in the ARCS and ANDS Business Plans.
- The Project Committee will meet quarterly via phone and/or agreed electronic medium.

Quality assurance processes to be used by or applied to the project

The ARCS/ANDS agreed arrangements will apply.

List of names against key governance and project management roles

The Project Committee is still being decided. Proposed members include:

- ARCS Executive Director, Professor Anthony Williams, or nominee
- ANDS Executive Director, or nominee
- Dr Allan Jones, Chair of AMMRF eResearch Committee
- Nick Hauser, ANSTO
- Richard Farnsworth, Australian Synchrotron
- Dr Peter Turner, Manager of the ARC Molecular and Materials Structure Network (MMSN)
- Director of VeRSI, or nominee

The Project Manager will be appointed by the Project Committee and specified in the Project Plan.

11.8.5 Project Summary

Deliverables / Milestones

The deliverables for the project are still under discussion. Some proposed deliverables are presented below but these have not yet been agreed. Details of deliverables and milestones will be specified in the Project Plan.

2008/09

- Investigation of possible solutions for the federated data repositories, metadata catalog and associated web portal, including work done by STFC, ARCHER, VerSI and GRANI project
- Investigate how the proposed system would interface to the ARCS data fabric and data storage and repositories at the MARCS and other organizations such as VerSI and universities
- Develop a design for the system building on existing software and the ARCS data fabric
- Design and prototyping of software and services for data and metadata capture from the various instruments and transfer to data repositories
- Development of software for converting data and metadata to standard formats
- Investigation of data sharing, and authorization and authentication mechanisms and development of prototype AAA services based on AAF
- Investigate approaches for automating workflows for data and metadata capture, conversion to standard formats and schemas, other data processing, and ingestion into data repositories.
- Deployment of services for data transfer from ANSTO and some Australian Synchrotron beamlines, AMMRF facilities and X-ray labs to MARCS and other organizations and institutions which have appropriate storage support.
- Functional data repositories for ANSTO and some AMMRF facilities.

2009/10

- Authentication and authorization mechanisms integrated with AAF and ARCS services.
- Development of documentation and researcher training material, some researcher training programs run
- Some data processing workflow services implemented
- A federated data repository and prototype collaborative work environment for AMMRF facilities
- Widening the deployment of the system to additional appropriate facilities, instruments, beamlines and data types not addressed in 2008/09
- Extension of services to interface other repository services such as provided by SPECTRa, e-Crystals and TARDIS.
- Ongoing investigation of interoperability issues and technology developments (e.g. iRODS)

Overall risk assessment

To be completed in the Project Plan.

Review points

Quarterly reviews by ANDS and ARCS, six monthly written reports from the Project Manager to the Project Committee, and a yearly review each September (starting 2009) by NeAT.

11.9 NeAT Round 2 Ideas that are being developed as possible proposals

These ideas are contained in a separate attachment, as many of the ideas were only received as PDF documents. A brief summary of each project is as follows:

- The Australian Node of the Human Variome Project: develop software, storage and protocols so that generators of data about human variation (genotype and associated phenotype) can easily deposit data for use by others in a de-identified mode.
- Australian Parameter Estimation Service: provide access to a range of HPC facilities dedicated to parameter estimation, a user centric help desk that assists researchers in porting their code to such a facility, and a software system that supports a range of parameter estimation techniques, based on the Nimrod family of tools.
- Emergency Animal Disease Bioresponse Collaboration Platform: provide an advanced, high capability telecollaboration platform (and its associated tools and services) to dramatically increase responsiveness to problematic biosecurity threats by its ability to quickly, and fluidly, bring in shared, visual, up to the minute information and data resources from a variety of sources.
- Transforming performing arts eResearch: apply collaborative networking, data visualization and mobile delivery to revolutionise the research utility of the AusStage database by providing performing arts researchers with platform-independent, remotely accessible and visually interactive access to quantifiable research data.
- Large Scale Temporal Spatial Ecosystem Digital Information Service: establish a federated, national criminal justice research data network that will provide a streamlined, single-point access to de-identified, unit-record level research dataset(s).
- Remote Computed Tomography reconstruction and visualization service at the Australian Synchrotron: a software service including a system for rapid Computed Tomography (CT) reconstruction and visualization of 3D structures of samples from the data collected at the Australian Synchrotron, tools for remote access to the reconstruction software via Internet and tools for remote visualization and collaboration capable of handling large images over Internet via network connections with realistic bandwidths
- Possible Datamining proposal: A number of ideas were submitted in the data mining space; a process is currently underway to determine if there is a possible service offering.
- Possible Humanities and Social Science collaboration environment proposal: A number of ideas were submitted in the HASS collaboration space; a process is currently underway to determine if there is a possible service offering.