



**Australian Government**

**Department of Education, Science and Training**

National Collaborative Research Infrastructure Strategy

Platforms for Collaboration

Australian National Data Service

DISCUSSION PAPER

May 2007

## PURPOSE

The purpose of this paper is to provide a platform for discussions to progress the Australian National Data Service (ANDS).

The discussions will take place at a Workshop to be held on Tuesday, 29 May 2007. The agenda for the workshop is at Attachment A.

In summary, the outcome sought from the workshop is a general consensus model of:

- the services that ANDS will offer;
- governance and implementation issues that will need to be addressed; and
- the process for determining who should be involved in the development and implementation of ANDS and its services.

## BACKGROUND

The proposal to develop the ANDS was included in the Investment Plan for the Platforms for Collaboration (PfC) capability of the National Collaborative Research Infrastructure Strategy (NCRIS). The Investment Plan was developed in consultation with the research community, and was approved by the NCRIS Committee following its April 2007 meeting. As well as the NCRIS Principles<sup>1</sup>, the Investment Plan drew on a specifically developed set of Investment Principles<sup>2</sup> (Attachment B).

The vision for the Data Management Infrastructure Component of Platforms for Collaboration is:

*All research data is appropriately curated and retained, and coordinated services support easy location, access to, and analysis of, that data.*

The Investment Plan sets out that ANDS will deliver:

- research data registration, location and access services,
- outreach services for researchers and institutions,

that can enhance the effective use of data within a federated research data management system, and

- stewardship of some nationally significant data collections.

A number of research projects, including many supported by the Systemic Infrastructure Initiative (SII), have addressed issues in these areas.

A key issue for ANDS is the successful transitioning of the outputs of these projects into an effective, fully operational service with a well defined quality of services that meets the needs of Australian researchers.

At the same time, the National e-Research Architecture Taskforce (NEAT) is being put in place to determine future infrastructure needs. Research and development activities, including new service/ system development, across all elements will be driven by NEAT priorities.

---

<sup>1</sup> <http://www.ncris.dest.gov.au/faq.htm>

<sup>2</sup> <http://www.ncris.dest.gov.au/NR/rdonlyres/2DFC56E4-83B4-4CFC-BC74-1CA4145A2657/16382/PfCInvestmentPlanFinalwebversion1.pdf>

## ELEMENTS OF ANDS

The Investment Plan grouped ANDS services into three main categories: Federated Services, Outreach Services and Stewardship Services.<sup>3</sup>

The experience of ANDS for an end user should be consistent across a range of services, regardless of the actual organisation delivering each service component.

In addition to the services set out below, ANDS will identify ways to assist researchers in their own data management and in the integration of their workflows with ANDS national data services and collections.

## FEDERATED SERVICES

The Investment Plan identified the following priority services and infrastructure elements as being able to be initiated very quickly:

1. a registry of research collections, including access services to those collections.  
Registry of research collections:
  - based on Online Research Collections Australia (ORCA).Access services:
  - for fully public data, existing search and access procedures could be adopted
  - for restricted data, this element of ANDS will need to work with the Australian Access Federation (AAF) to provide secure access to data where required.
2. national discovery services across the network of Australian digital research repositories and the collection registry, eg.
  - National Library Resource Discovery Service
  - OAI harvesting/ search service for general text data
  - discipline-specific services such as the ANZLIC Australian Spatial Data Directory
3. persistent identifier services for digital objects
  - building on the work of the PILIN project; includes local identifier services and utility services to support uniqueness, resolution to the current location of the item and management of naming authorities
4. services for data re-positioning between repositories, for performance, policy and privacy reasons
  - there are elements of this work underway in DART and APSR as well as the APAC Data Grid
5. a metadata and ontology schema registry to assist interoperability and discoverability of data and encourage standards, including
  - building on the work of DART, ARCHER and the Demetrius repository at ANU
  - using the expertise in the USQ standards team

Once these support services are established, common data analysis and visualisation services, generic data quality assurance services, notification and data curation services, and common data submission and presentation services can be layered on top.

Future activities should include cross-discipline terminology services, aggregated statistics service, workflow services.

---

<sup>3</sup> The relevant excerpt from the Investment Plan is at Attachment C of this paper (pp.10- 14)

## OUTREACH SERVICES

There are two broad types of services within the category of outreach services:

### Strategic

- Activities oriented to the institution or key stakeholder relationship
- Setting, promoting and evolving ANDS integration requirements
- Assuring institutions that their approach to data management and associated investments will be sustainable, and providing advice where they are not
- Community building to assist with making data management sustainable, including promoting the development of communities of practice based in disciplines that are not currently well organised, to share and embed good practice.
- Ideally, would involve some cross-over of people from institutions on major data system acquisitions and institutional planning.

Of the specific outreach services set out in the Investment Plan, the following can be categorised as strategic:

1. create networks of expertise to bring together researchers and research technologists to develop recommendations against data management issues
  - draw on expertise currently existing in many SII-funded projects eg. MAMS, APSR, RUBRIC, APAC. Build on network/s of NCRIS capabilities.
2. provide high-level consultancy
  - Focus would be on embedding good data management practice in organisations/ disciplines that don't have it but could benefit from it.
3. liaise with other appropriate government and non-government data initiatives, such as the ABS National data network, and water, geospatial and health data networks
  - with the aim of achieving consistent good practice in data management and greater interoperability of systems and data.
4. collect and disseminate information on best practice in e-Research data management policy
  - a range of research reports currently exists in the area of data management, data collections, curation and good practice. Finding ways to have their key messages distilled then disseminated widely is the challenge here.

### Tactical

- Helping communities, including small groups of researchers, assess their options and develop solutions to their needs, in the context of the options available to them at their institution or more broadly (following on from the strategic outreach)
- Provide a 'hand-holding' rather than 'data-holding' service. Could include intensive, one-on-one assistance

As opposed to the strategic element, it would make more sense for this 'tactical' component to be delivered by local providers.

Of the specific outreach services set out in the Investment Plan, the following can be categorised as tactical:

1. respond to a strategic agenda around skills development and awareness building for research data in Australia by holding appropriate events, workshops and seminars

- this could be achieved by a range of possible approaches, including state- or regional- based outreach activities, drawing on the PACs; the APSR outreach model; other possible/ hybrid approaches.
2. co-ordinate distributed sources of expertise to support individuals and projects develop improved data management practice
    - this service would support not only discipline-specific researchers but also other groups such as libraries, research offices and IT Services within universities to improve data management practice.
  3. develop data architectures for specific communities and multi-disciplinary research projects
    - could connect here with the Interoperation and Collaboration Infrastructure (ICI) Component/ compute tools and discipline services to propose activities within that program.

There is a range of issues that will arise in relation to tactical outreach:

- What if researchers/ communities respond: 'if you want to tell us what to do with our data, why don't you take it?' ANDS does not have sufficient funding, or a mission, to hold a great deal of data. Specific funding is being provided to institutions from multiple sources to provide infrastructure that could be extended to support retention of data (eg. ASHER).
- It may be that researchers are protective of their data and concerned about how it can be safeguarded in the institutional context.
- How to resource and motivate the service?
- The criteria and methodology for identifying the priority targets. These kinds of services will need local people on the ground. ANDS could take an Expression of Interest (EOI) approach to determining priorities.

## **STEWARDSHIP SERVICES**

The stewardship activity aims to support continuity of access to significant public collections of data. Some of the activity categorised as stewardship in this document and in the Investment Plan will also be applicable as outreach services at an institutional or discipline level.

This activity includes the following main components:

1. curation and collection management services, covering areas such as
  - a. metadata management, including advice on use cases and content models
  - b. curation
  - c. preservation
  - d. archival
  - e. legal and licensing.
2. facilitating, co-ordinating and enabling a national network of digital repositories, through the provision of a consultancy, coordination and advisory service.

In addition to the list above, a further possible stewardship service has been identified:

Collect and disseminate information on best practice in use of repositories to support the collaboration-publication continuum.

- this would build on the 'life-cycle of research' work that the ARROW and DART projects in particular have been doing, and reflects the thinking of the JISC RIN report on data, that it is necessary to look at both publications and data as different parts of the same problem.

Stewardship services might include some limited provision of shared data storage (physical infrastructure) but only for those collections without a natural home.

- A small number of organisations could provide large scale shared data storage facilities to ANDS to support the hosting of and access to significant research data collections. This will include collections from a range of disciplines and in some cases will need to be co-located with compute capacity.

ANDS may also assist in brokering storage, assisting data generators to find the appropriate storage facilities or assisting in negotiating storage options.

## **GOVERNANCE**

DEST has considered a range of options for the governance of ANDS. Delivery of the strategic outreach element and coordination of the tactical element have to be consistent with ANDS and Government policy and delivered coherently on a national basis.

It is therefore DEST's view that there is a real requirement for a central entity that fulfils some but not all of the important functions of ANDS. This would be complemented by a number of 'satellite' entities/ service providers, whose role is outlined below.

There is some more work to do about defining in detail what functions are in the centre and what are able to be delivered by the satellites. For example, aspects of standards development and implementation and the management of access to data might be located in both.

### **Role of the 'central' entity**

The central ANDS entity will have a key role in setting the high-level agenda for ANDS.

It will:

- manage and operate Federated Services
- incorporate work on establishing standards for interoperability
- conduct 'strategic' outreach
- fund and coordinate 'tactical' outreach activities.

It will also have a role in facilitating cross-community linkages.

### **Role of the 'satellite' entities/ service providers**

The 'satellite' entities will provide ANDS services through a variety of arrangements. It is expected that they will be located within existing organisations or institutions. In some cases, they may service a particular community/ discipline or capability. These services will be provided on a nationally agreed basis, in line with NCRIS principles.

A key role for these 'satellite' entities will be the stewardship of data, for one or a range of connected community/ discipline or capability groupings, or for some that are not currently well organised. There will be an emphasis, connected with the stewardship role, on 'tactical' outreach to give research communities the skills and tools they need to manage their data.

Organisations prepared to offer ANDS services, including tactical outreach, will do so under the ANDS 'brand.'

The ANDS website will have a list of organisations that provide ANDS services, but end users will approach ANDS itself as the overarching organisation that provides access to data management and federation services.

## WAY FORWARD

Following the workshop, a small working group will be formed to advise DEST on how to proceed in implementing the outcomes of the workshop.

The next phase of the work in defining ANDS should get us to a point where we could call for Expressions of Interest to deliver ANDS functions, including the central entity role.

The central entity could be a single institution or organisation, but it could just as appropriately be a small consortium.

There are a number of principles that should inform the nomination and ultimate selection of an institution or consortium to be the central ANDS entity:

- leadership in the field of data management
- appropriate experience in the provision of services
- agreement on a single “head office” for ANDS - not enough funding to support more than one.

While it appears attractive to prefer a consortium of several institutions, they would have to advise DEST as to how the consortium would achieve the requirements.

It is also preferable that any consortium be put together by the proponents rather than brokered by DEST. Ideally, there would be one agreed response to the EOI for the central role, and a number of complementary responses for the provision of satellite services.

It should be noted that while DEST will seek to implement the ANDS model developed at the workshop, it retains discretion over NCRIS funding until an effective and suitable governance mechanism is identified and implemented for ANDS.

The research communities that ANDS seeks to support will be kept informed, and appropriately involved, in the process.

National Collaborative Research Infrastructure Strategy  
Platforms for Collaboration

**Australian National Data Service (ANDS)  
WORKSHOP**

Agenda

1. Welcome and Introduction  
Dr Evan Arthur, Group Manager, Innovation and Research Systems Group, DEST

Workshop Facilitator

Professor Mick Reid, consultant and former Director-General of Health for NSW

2. What elements should be a part of ANDS?  
Which of the services are of the highest or most immediate priorities?
  - i. Federated Services
  - ii. Outreach Services
    - a. Strategic
    - b. Tactical
  - iii. Stewardship Services
3. Discussion of proposed governance model
4. Feedback on proposed EOI process
5. Next Steps

## Key Principles for investment in research infrastructure under NCRIS

The key principles underpinning NCRIS are that:

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

## Platforms for Collaboration Investment Principles

Developments in e-Research and cyber-infrastructure, both technological and social, are rapid and likely to continue for the foreseeable future. Hence, we can expect an increasing pace of change and an ongoing flow of new opportunities to enhance the quantity, quality and productivity of research efforts; noting always that some research is otherwise impossible and that the improvement in infrastructure and the ability to ask more demanding questions go hand in hand.

Combined with this pace of change, the independent goals and decision making processes of a multitude of governments and institutions will ensure that any future Australian e-Research infrastructure will be a combination of many activities with a need for significant co-ordination.

*Therefore: A foundation principle is to enhance, strengthen and build on co-operative arrangements so that an increasingly coherent level of support can be provided to researchers, and their collaborations and communities. (A)*

The strengthening of e-Research activity and the pursuit of collaborative research within existing ICT intensive disciplines and more broadly within other disciplines hinges on the ease with which e-Research activities can be carried out.

*Therefore: A foundation principle is to reduce barriers to adoption of e-Research by systematically supporting new e-Researchers as well as expert e-Researchers; by sourcing and supporting suitable tools and services. (B)*

e-Research can most usefully be contemplated as a permanent change in the way researchers work; so that the capabilities needed to carry out e-Research must be robust and enduring, reliable, always on, and commoditised when practical.

*Therefore: A foundation principle is to provide robust and enduring services delivered by providers where the delivery of the service is the mission. (C)*

The funding available from the National Collaborative Research Infrastructure Strategy (NCRIS) for Platforms for Collaboration (PfC) is significantly less than would be required to support all the Australian research which could benefit from e-Research services, so a means of defining priority is required.

*Therefore: A foundation principle is to prefer e-Research infrastructure services which are of value across multiple research communities; and are of value to those communities whose research needs led to the NCRIS capabilities. (D)*

## The Data Management Infrastructure Component

This component will bring together data management interests to work towards the vision:

*All research data is appropriately curated and retained, and co-ordinated services support easy location, access to, and analysis of, that data.*

The component is intended to be a step towards a visionary landmark research infrastructure that meets Australia's future research data needs.

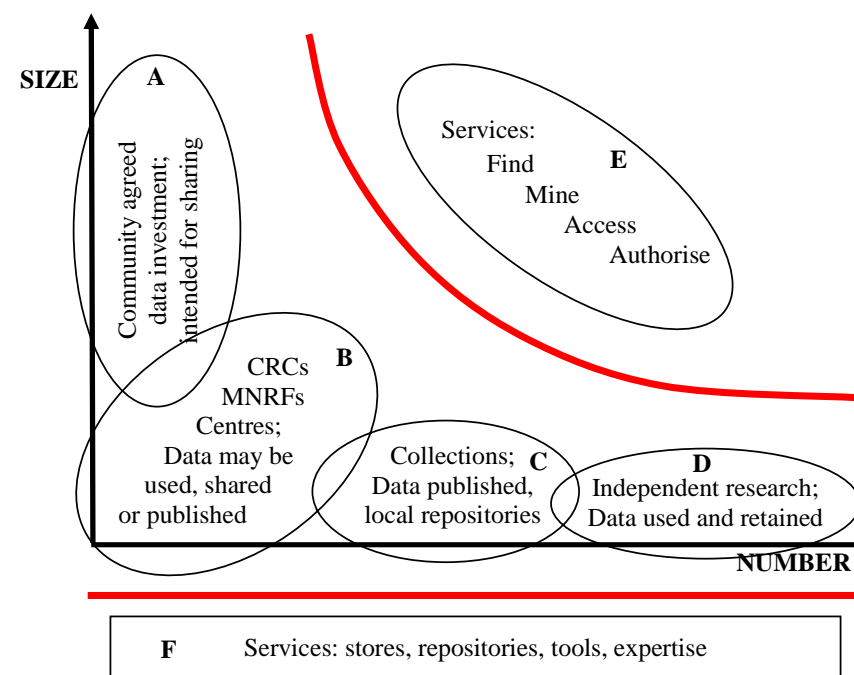
To create momentum in that direction, the component will implement a strategy to help experts and the builders and users of curation, access and analysis tools to work together to improve community data management practice — to 'keep data better'. It will provide:

- Outreach advisory and support services to assist researchers establish data curation practices
- Outreach advisory and support services to assist institutions establish research data management infrastructure and practice in a nationally coherent manner
- Federation level services, including data location, access and data mining
- Data management brokering to help match needs to providers
- Stewardship services hosting some categories of collections
- A community based merit allocation scheme for managed resources

Consultations and surveys undertaken by Platforms for Collaboration suggest that data volumes are growing rapidly and that much research data is held in isolated forms and not easily accessible.

Generally, two kinds of data can be distinguished. Private data is often held by a researcher through self-generation, or by acquisition from surveys, an instrument or device. Public data refers to data placed in a public repository for general access.

Often, researchers also make some private data available on a controlled access basis, but do not normally provide public access to all the data they hold. Also, where data is derived from significant public or shared investment, a common data holding protocol often applies where researchers are granted controlled access to the data collected on their behalf for a limited time. After that time the data reverts to the researcher's host institution, or is made public through a repository.



With reference to the diagram above, in developing a rationale for NCRIS investment, the following missions, properties and responsibilities seem to be present.

- (A) Several communities capture and share data of general value, where large datasets are generated or gathered as the result of significant investments; such as in astronomy, high energy physics, earth observations and bioinformatics. The size of these data sets tends to be very large, typically in terms of tens to hundreds of terabytes and sometimes petabytes, with high growth rates.
- (B) More frequently, organisations, such as Co-operative Research Centres (CRCs) and Major National Research Facilities (MNRFs), have data gathering, generation, curation and publication as part of their mission; or use data for internal purposes, the results of which are published as an information service; or they provide research support services that generate data. The data often is privately held and only processed information is published. The size of these data sets tends to be in the many gigabytes to terabyte range.

- (C) Even more frequently, many research organisations, departments, teams, and even individuals have established data collections the contents of which are intended for publication and access. These publicly available data sets (donated by researchers) are often kept in institutional or personal archives or web sites. The size of these data sets tends to be in the multiple gigabyte range.
- (D) Nearly all researchers generate or store data on the desktop, much of which is only ever intended for individual use and which is inaccessible to other researchers, visible only by publication of derivative research results. Most of this raw data is intrinsically hard to re-use and could only be accessed if it was moved off the desktop. The size of such data sets is highly variable.
- (E) Some organisations provide access to research data, either within disciplines or more broadly, and may or may not hold the primary data themselves. These organisations provide catalogues, abstracts or thumbnails, and can search and possibly mine data over which they hold indexes. These sources can only easily provide access to public data. The aggregate size of data accessible in this way would be in the range of terabytes to petabytes.
- (F) A final class of mission relates to the retention, curation, access and analysis of data sets. Generic service providers necessarily operate only on public data, although some specialised services could operate on controlled access data. The co-location of data sets for data mining and linking is often part of the value of these services. Many institutions might operate such a service on their own behalf for their retained private data. The size of repositories tends to be in the range of hundreds of terabytes to many petabytes, and exabytes would be ultimately conceivable in some settings.

As the discussion suggests, data is everywhere, and the aggregate investment across

all of the missions and user communities is large and rising quickly.

It also shows that data management services will be provided by a variety of sources and investment by NCRIS needs to be based on principles that fit within a broader framework. The first two columns in the table below are derived from policy under development by the Australian Vice Chancellors Committee (AVCC), National Health and Medical Research Council (NH&MRC) and the Australian Research Council (ARC) and the third summarises the investments developed here.

The investment proposal was also informed by PMSEIC, which noted:

“The Data for Science Working Group discussed at length the idea of a new National Centre for Data for Science. There was considerable support within the Group for a Centre; it was felt that such an initiative would be of benefit and may be a useful mechanism for progressing many of the above recommendations.

The Working Group considers that there is a range of functions that a Centre could assist with, including:

- facilitating and promoting the changes reflected in the recommendations;
- working with those in specialist scientific disciplines to discover datasets;
- establishing vital repositories; and
- working collaboratively with the research, government and business communities to support the proposed new approaches to data for science.

“The Working Group stopped short of recommending the establishment of a Centre. The Working Group concluded that the high-level expert committee (recommendation 2) should decide whether such a centre was desirable and, if so, where it may be hosted, and what its role and governance mechanisms should be.”

<p>Institutions should:</p> <ul style="list-style-type: none"> <li>Develop and implement a policy on data ownership</li> <li>Provide guidelines to researchers on ownership, what to keep and researcher responsibilities</li> <li>Maintain durable records on what research data has been held and ensure that research data is under the control of the institution where the work was performed</li> <li>Provide secure systems for holding data and for granting access to that data.</li> </ul>	<p>Researchers should:</p> <ul style="list-style-type: none"> <li>Determine what data to keep, considering research community practice and any project or legal requirements</li> <li>Ensure research data is retained (for at least 5 years from publication of results) using institutionally provided mechanisms</li> <li>Ensure at the end of employment (for whatever reason) data retention passes to the institution</li> <li>Maintain confidentiality where it exists.</li> </ul>	<p>NCRIS should:</p> <ul style="list-style-type: none"> <li>Provide national implementations of the federated services described under categories E and F</li> <li>Build expertise and provide outreach services that can assist others, including ‘training the trainers’</li> <li>Broker solutions for collections and researcher needs and support identified collections to fast track e-Research development where appropriate</li> <li>Ensure promulgation and use of agreed legal frameworks and templates governing access to data.</li> </ul>
--	---	--

## Implementation Strategy

The Platforms for Collaboration investment will establish an Australian National Data Service (ANDS) as a cooperative centre with expertise in research data management.

ANDS will address the needs raised by PMSEIC by providing three mutually reinforcing services focussing on shared services, stewardship and outreach.

While it would be possible to focus on any one of these and still provide value, the development of a national centre of expertise will be significantly enhanced by embodying and bringing together knowledge from all three.

With reference to the previous analysis of the two meta-service categories depicted in the diagram, those related to re-use (E) and those related to retention (F), data search, data mining and access control services in (E) and national data location and movement services in (F) are systemic, beyond the bounds of individual institutions, and complementary and value adding to individual data holdings.

ANDS should therefore focus on providing common services in support of research data collections and provide integration infrastructure that facilitates sharing of data, so that researchers can more easily discover, access, use, analyse, and combine digital resources as part of their activities.

While simple services will come first, the longer term vision is to connect Australian and international data centres, repositories and online collections to enable serendipitous discovery, cross-disciplinary research, and cross-repository workflows and to transform the disparate collections of research data around Australia into a cohesive corpus of research resources.

ANDS should also provide specific services that support collections of public data (starting with collections in category (B) and (C)) that could be preserved and made accessible. This service adds considerable value to the data and will not otherwise be easily or quickly provided. Over time, specific services may grow or diminish as research institutions establish a position on research data management.

ANDS should also be heavily involved in the identification, installation and adoption of user-centric tools, and the engagement of the research community and stakeholders through consultation and outreach activities. ANDS personnel involved in outreach will also be grounded in day-to-day data management and data services (and vice-versa). In general, ANDS should not set up new infrastructure facilities, but rather extend and build-on existing or proposed capabilities.

The development of ANDS will need to proceed in a manner that allows potential participants to understand the proposed functionality and negotiate their involvement.

AeRIC should ensure that:

- an interim ANDS Governing Committee is established to provide guidance on the further development of ANDS;
- a clear statement of the goals of the investment in ANDS and the functionality it is expected to create is made available to potential participants;
- expressions of interest in ANDS are invited from potential participants, in consultation with the interim Governing Committee; and
- an ANDS NCRIS agreement and initial business plan are developed in consultation with participants and recommended to NCRIS by December 2007.

ANDS should be empowered to:

- receive funds from AeRIC under the arrangements defined in the NCRIS agreement and against the purposes and activities identified in the business plan;
- operate as a single co-operative activity across a network of contributing partners allocating aggregate resources against needs through an agreed merit process;
- develop a community of interest around research data management, and engage in relevant international developments;
- develop and sustain the operation of federated services that enhance access to research data;
- develop and sustain an outreach activity that assists researchers and research organisations to improve their research data management practice;
- develop relationships with custodial service providers to deliver an effective national collections management service;
- appoint the ANDS Director;
- manage resources provided to ANDS within the terms of the business plan;
- report to AeRIC on the implementation of the business plan on a regular basis and
- provide a progress report and propose a revised business plan on an annual basis.

## Federated Services

The establishment of a set of common services over a national network of digital repositories is the primary goal of ANDS. While those services will be defined and co-ordinated by ANDS, they will be implemented by appropriate service providers.

In the short term, the following priority services and infrastructure elements can be initiated very quickly:

- a registry of research collections, including access services to those collections
- national discovery services across the network of digital repositories and the collection registry
- persistent identifier services for digital objects
- data re-positioning services between repositories for performance, policy and privacy reasons
- a metadata and ontology schema registry to assist interoperability of data and encourage standards.

Expertise around these services already exists in Australia, with work underway. The intention is to adopt, adapt and influence this existing work to create the ANDS services and to engage in international co-development where appropriate.

Once these support services are established, common data analysis and visualisation services, generic data quality assurance services, notification and data curation services, and common data submission and presentation services can be layered on top.

Future activities should include cross-discipline terminology services, aggregated statistics service, workflow services, etc.

ANDS will also identify tools that assist researchers in their own data management and in the integration of their workflows with ANDS national data services and collections.

## Outreach Services

This activity aims to improve the expertise, capabilities, and most importantly, organizational policy and everyday practice related to research data production, management, structuring, description, analysis, and curation.

The intention is to support researchers, data scientists, and IT infrastructure staff understand the developing requirements around research data management in order to participate in this new age of digital collections; and to assist managers gain confidence in the suitable policy options that they might adopt.

As a result, the ANDS outreach will help develop a national research data management framework and networks of expertise by bringing together researchers and research technologists to develop recommendations against data management issues. It will:

- respond to a strategic agenda around skills development and awareness building for research data in Australia by convening appropriate events, workshops and seminars
- co-ordinate distributed sources of expertise to support individuals and projects develop improved data management practice
- provide high-level consultancy
- develop data architectures for specific communities and multi-disciplinary research projects
- liaise with other government data initiatives, such as the ABS National data network, and water, geospatial and health data networks
- collect and disseminate information on best practice in e-Research data management policy.

## Stewardship Services

The Stewardship activity aims to support continuity of access to significant public collections of data.

This activity includes three main components:

- curation and collection management services, covering areas such as metadata management, curation and preservation, archival, legal and licensing, etc.
- facilitating, co-ordinating and enabling a national network of digital repositories
- limited provision of shared data storage (physical infrastructure) but only for those collections without a natural home

It will be distributed across a federation of organizations with an intrinsic interest and expertise in data management. NCRIS funds will extend those services in ways that meet the ANDS objectives.

In providing data storage, ANDS will only provide co-location and access methods for community-level collections which occupy the middle-ground between national reference collections and individual researcher collections. Ongoing support for such collections is often beyond the reach of a single institution and below the threshold that could attract recurrent support at a national level. Pre-requisites for a data collection to be supported by ANDS would include: a high level of community engagement, high quality structured data, adherence to international standards, and the on-going participation in data curation activities by the community. In general, ANDS will not fund generic data hosting services.

Curation activity will be a shared responsibility between the ANDS and identified research communities.

## Next Steps

The data community is large and diverse, and while support exists for the focus taken in the Pfc planning, the alignment of participants needs considerably more work.

The proposed path forward is to clearly articulate the requirements Pfc seeks to support through infrastructure investment; and engage in a highly intensive further consultative process to arrive at participants and contributions towards those goals.

An overall envelope of NCRIS funds retained for ANDS is set at \$20M and an indicative outline of funding against ANDS activities budget is indicated below.

	NCRIS	Other	Total
Federated services	\$M	\$M	\$M
Operational	2	0	2
Developmental (NEAT)	5	10	15
Stewardship	7	7	14
Outreach	7	7	14
TOTAL	21	24	45

### Stewardship

A level of funding will be provided, increasing over time, through which ANDS will be able to assign support for collections.

The detailed estimates are difficult, as different data and different access regimes demand different service support levels and associated costs.

For instance, some collections will need some curation work as part of their acceptance; some data will require information services to be hosted; some data will require higher access rates and a larger proportion of underlying disk to tape support; and some data will require off-site replication and so on.

The approach is that the funding will remain relatively set and the capacity available for collections will depend on the requirements of the holdings as they are accepted.

The rising funding line and the trend in storage system capabilities will ensure that a significant number of new collections can be accepted each year.

ANDS will have a range of capabilities in its retention providers so that the needs for different collections can be met by appropriately configured and costed services.

ANDS will also provide a brokerage function to assist other collections identify providers most related to their area of interest.

### Federated services

The scale of cost associated with the federated data services is more related to their development and extension over time than their annual operation.

While this could be provided in activities in other programs, it is placed in ANDS to ensure that governance arrangements around ANDS can determine the priority for the tools and services to be developed.

As ANDS becomes more operational, this funding level will decline, so that the more advanced location and analysis services are more user community determined.

The ANDS program will, however, support the installation and operation of servers and the configuration and maintenance of the services when operational.

Some of the resources for development can be provided by the ICI participants in order to ensure a close integration with operational aspects of target services and to simply transition into production support.

Development activities must be driven by infrastructure needs determined by NEAT.

### Outreach

The outreach program will be funded on an in-kind co-investment basis with organisations that also seek to address the curation difficulties of their local research communities.

Some support will be provided in every region.

ANDS will work to identify communities to which curation assistance should be provided for improving curation or data management practices, and to also identify collections to be accepted into the stewardship activity.

The program agreement will provide that a significant portion of this effort is to work on collections identified for stewardship by ANDS.