

## **NCRIS – Australian National Data Service (ANDS) – AARNet Statement of Capability**

AARNet Pty Ltd is the not-for-profit company that owns and operates one of the world's most advanced and wide reaching national and international research and education networks. With more than 30,000kms of fibre assets nationally, AARNet has the largest footprint of any national research and education network in the world. Our vision is a high bandwidth Australian Research and Education Network, in which the tyranny of distance does not impede Australian research, teaching and learning collaboration both locally and internationally.

Our strategy is to capitalise on our investment in long term infrastructure with network innovation for the benefit of future generations of Australian researchers, educators and students. Our focus is our dedication to research and education and connectivity to the other advanced international and national research and education networks around the world.

AARNet provides high-capacity communications services to its Shareholders, the 38 Australian universities and CSIRO, as well as to other research and education institutions in Australia and the South Pacific. Managing the network is our primary function. AARNet operates a 24-hour, 7-day response service, achieving extremely high levels of service on both the network and other services. The differentiator is many gigabits of bandwidth with no congestion both nationally and internationally.

AARNet generally connects customers via customer edge equipment to ensure the integrity and performance of the network. However, where required AARNet connects and supports directly-connected servers such as the MAMS Testbed Shibboleth Federation, the RETAIN file store which acts as a Mirror and provides other network-connected storage capabilities.

AARNet has a staff contingent of 38 with a significant presence in every capital city except Hobart. Against this background, AARNet can offer the following capabilities to ANDS:

1. Connection to the AARNet3 network.
2. Specialised network connections (e.g. dedicated light paths) between 2 or more sites.
3. Provisioning of special circuits for experimental use in research.
4. Provisioning "private" networks among closed communities – "research commons", ad hoc or lasting.
5. Civil works for connections into the core network.
6. Videoconference and VoIP services operating over the network.
7. Connection to the network of devices requiring high-availability and high-performance connections.
8. Location and management of specialised servers, including for instance a file store which could be used to aggregate the smaller storage needs of organisations who find it uneconomical to provide one themselves.
9. Skilled technical staff in most capital cities to participate locally in projects requiring high network connectivity.
10. An Operations Response Centre that could be expanded to cover other services of national importance.
11. An incorporated entity and neutral aggregation point with a proven willingness and ability to collaborate with a wide range of users across the research and education community, nationally and internationally.
12. A lightweight governance structure – AARNet already has contractual agreements in place for connection to the network with all of the institutions in the higher education and research sector throughout Australia.

TAR: 7-Jun-07

## National Data Network (NDN) Consortium - ANDS Capability Statement

The NDN Consortium [ABS, AIHW, ANZLIC, ARACY, Centrelink, CSIRO, DEST, DoHA, DoTaRS, Office of Spatial Data Management, Office of the Privacy Commissioner, Telethon Institute of Child Health Research, NSW Health, QLD Treasury, VIC CIO, SA Cabinet Office] was established to develop a distributed network of information resources to improve visibility, accessibility and use/re-use of information resources in the Survey and Administrative Data Domain. ABS leadership of the Consortium arises from its role as the national statistical agency (since 1905) and its legislated responsibility "to ensure co-ordination of the operations of official bodies in the collection, compilation and dissemination of statistics and related information" (Under s6(1)(c) of the ABS Act 1975).

The NDN:

has been governed by an Interim Governing Board comprised of senior management representatives of Consortium member organisations

architecture, while developed with a particular data domain in mind, can serve other data domains equally well.

is an open sources development, that has undergone 2.5 years of systems development, involving collaboration with Macquarie University (Shibboleth), Queensland University of Technology (Creative Commons) and the US Census Bureau (DataFerrett)

has been in a demonstration phase since 1 July 2006 and is entering a 12 month pilot phase on 1 July 2007

Further information on the NDN can be found in:

the PMSEIC 'Data for Science' report presented on 8/12/06:  
[http://www.dest.gov.au/sectors/science\\_innovation/science\\_agencies\\_committees/prime\\_ministers\\_science\\_engineering\\_innovation\\_council/meetings/sixteenth\\_meeting.htm](http://www.dest.gov.au/sectors/science_innovation/science_agencies_committees/prime_ministers_science_engineering_innovation_council/meetings/sixteenth_meeting.htm)

the NDN website: [www.nationaldatanetwork.org](http://www.nationaldatanetwork.org).

At its meeting of 30 May 2007, the NDN Interim Governing Board endorsed the NDN consortium developing ANDS proposals that build on NDN infrastructure, and that unify future development and operation of ANDS/NDN facilities.

The NDN Consortium has an extensive range of capabilities it can draw on from member organisations. At the same time, given ANDS requirements go beyond the Survey and Administrative Data Domain, the Consortium is interested in other organisations joining the Consortium in the development of ANDS proposals.

ooo000000000ooo

## What ANDS might look like

### **Components of ANDS**

ANDS covers four broad areas:

1. physical storage of data;
2. development of a system for overall national data management;
3. operations of ANDS;
4. providing end user support to researchers.

### **Physical storage of data**

ANDS should not care where data is physically stored, as the overarching management system will be aware of where data is.

There are quite a few sizeable silos for research data around the country at present (ACT, Vic, QLD), but none in NSW, unfortunately. It is not clear whether ANDS will fund physical data storage, as there are competing priorities, but if funding is available through ANDS, this would provide an opportunity for the acquisition of a sizeable data storage facility, which would be located at **ac3**, adjacent to the supercomputers that would invariably be used to analyse the data.

### **Development of the ANDS Management framework**

This is a project — i.e. it has a start point and an end point, although there will invariably be ongoing enhancements, like any product. The specifications are yet to be developed, but there will be two main functions:

1. **Depositing data:** When a research project is being wound up, or when researchers wish their results to be available globally, they submit their results to ANDS via an intuitive web-form. This web form captures all the necessary meta data, which is used to create the directory entry for the data. If not already done so, the data is then ingested to one of the ANDS approved data repositories. This activity happens once only.
2. **Retrieving Data:** Researchers use this facility to source and retrieve data. Whilst the depositing of data occurs once, data retrieval may occur many times. A search engine is required, along with appropriate permissions engine (sourced from AAF). The CSIRO developed *FunnelBack* is a candidate search engine.

### **Operations of ANDS**

Having completed product development and beta testing, then the operational system needs to reside somewhere where there is 24x7 support. It is proposed that **ac3** host the ANDS infrastructure (which would be funded by NCRIS), providing first level support through the existing Support Desk. Second level support would be provided by Melcoe. The hardware required would not be significant — possibly a web server and a business server, both duplicated for redundancy — but they do need to be managed like any other operational system.

### **Supporting end users**

A research team has, to their delight, discovered several sources of data that relate to their current work, stored physically in different parts of the country, and perhaps overseas. ANDS informs them that they are entitled to use the data, perhaps with some restrictions. But the databases are all stored in different formats. So they approach one of the state-based e-Research Support centres for assistance. In due course the IT experts in the eResearch Support Centre provide a framework to the research team that make the disparate data bases look like a single logical database. The research team is able to carry out their research without the need to hire an IT specialist, or to become part time computer scientists themselves.

In the case of NSW, this end user support would be carried out by the NSW eResearch Support Centre.

P. McCrea

# ANU Capabilities for NCRIS Australian National Data Service

---

ANU has an established track record as a national leader in research data management, having acted as lead institution in a number of national initiatives, such as the Australian Partnership for Sustainable Repositories (APSR), the Australian Social Science Data Archive (ASSDA), and the Australian Partnership for Advanced Computing (APAC). ANU therefore has a long standing commitment not only to the importance of research data but also to *creating cohesion and providing focus within the sector*. ANU staff have successfully managed, delivered and coordinated national services for research data and worked within various governance frameworks. ANU also has a long history of supporting the national interest by the leveraging of ANU intellectual capital in forming national services in this field.

▪ **ANDS CENTRAL ENTITY:** ANU is able to **lead and manage the ANDS central entity** and work cooperatively with other parties as appropriate. ANU is ready to collaborate with other key players in the field to provide the services outlined in the ANDS Pfc investment plan.

▪ **DATA STEWARDSHIP:** Mature national data services at the ANU such as the [Mass Data Storage Service](#), the [Demetrius](#) Institutional Repository Service, [ASSDA](#), the [Australasian Pollen and Spore Atlas](#), and the [International Economic Databank](#) give ANU the capability to offer both **curatorial expertise** and **data hosting services** to ANDS. In particular ANU has provided national leadership in hosting and developing research support relationships with nationally significant data collections. Several landmark white papers and survey [reports](#) written by ANU staff (on secondment to APSR & APAC) provide a strategic background for data stewardship which has influenced national debate. ANU has built cohesion in this sector and would thus be able to catalyse development of **a broader national federation of research repositories** which will be key in the sustainability of the stewardship solution.

▪ **FEDERATED NATIONAL SERVICES:** ANU staff are developing the Online Research Collections Australia ([ORCA](#)) registry of research collections. ANU is keen to take the ANDS opportunity to continue supporting the technical and social development of this national **registry/ discovery infrastructure**. ANU will continue its alliance with the NLA, especially in the ongoing delivery of the related registry, discovery, and access services.

The ANU is committed to developing Australia's **data re-positioning infrastructure** and build on the work of the ANU SF and the APSR [Repository Interoperability Framework](#). For over a decade, the ANU has led the agreements and technology that support Australian research communities' data re-positioning needs.

The ANU has established high level expertise in providing many of the "later year" federated national services such as **data submission, presentation, fusion, and quality assurance** as well as **aggregated statistics, and obsolescence notification services**. [Existing](#) ANU projects, systems and services can form the basis of these federated national services.

ANU has a strong track record in the **strategic planning of national services for research data**. APSR and ANUSF have existing dedicated programs to develop national infrastructure for research collections.

▪ **OUTREACH:** ANU currently coordinates mature national and cross-organisational outreach programs for research information management, unique in Australia. The APSR program comprises [events](#), [publications](#), [training](#), [support networks](#), and consultations and spans both tactical and strategic concerns. ANU staff have experience in managing and delivering formal outreach programs, and ANU can build on these existing capabilities to **deliver ANDS strategic outreach and plan and coordinate ANDS tactical outreach services**.

▪ **SUMMARY:** ANU's existing experience and ongoing interest spans every aspect of ANDS:

	Central Entity	Data Stewardship	Federated Services	Outreach
Coordination	✓	✓	✓	✓
Service Delivery	✓	✓	✓	✓

## **AUSTRALIAN SOCIAL SCIENCE DATA ARCHIVE**

### **Expression of Interest to be a Satellite Entity**

The Australian Social Science Data Archive is a distributed network with nodes currently operating at ANU, UQ and UNSW - with UWA coming online in 2007. The team of professional data archivists, directed by a panel of Australia's leading social scientists, provides both stewardship and outreach services to the Australian Social Science community. ASSDA adopts and develops standards in line with international best practice and applies these standards across the distributed network. ASSDA currently provides a 'one-stop shop' for data acquisition and distribution via web-based services. In ASSDA's strategic plan (see Attachment A), it is expected that by 2012, each state will house at least one node of ASSDA provide further stewardship and outreach services to their local communities. Funding from NCRIS could increase the momentum of this process and support the branching out into stewardship and outreach services for the Humanities.

#### **Stewardship**

ASSDA has acquired, preserved and encouraged online access to social data including surveys, opinion polls, census, and is experimenting with curation methods for qualitative data (Australian Qualitative Archive – AQuA, based at the University of Queensland) and historical statistical publications (Historical Census and Colonial Data Archive - HCCDA). ASSDA regularly provides training to its archivists on metadata creation, standardisation of files, and the legal implications of the Privacy Act and the Copyright Act. An important new role for ASSDA lies in employing and training the data archivists needed to curate and manage the collections of these complex and varied data types. A professional cohort of data archivists is a major obstacle to furthering data curation and distribution services in Australia (especially compared with UK and US) and NCRIS potentially provides an opportunity to establish this profession beyond the social sciences.

ASSDA's distributed model has allowed the development of curation practices and metadata standards to address specific problematic data types and to experiment with curation, outreach and governance structures to support data existing in closed systems such as Indigenous data. With small amounts of establishment funding, new nodes can be developed to solve these problems, drawing on the strength of ASSDA's previous experience and have an instant network of project partners to support funding applications. With further resources, the ASSDA stewardship role could also spread to the humanities. Using standards developed through ASSDA's Qualitative Archive service to manage research recordings for text, video, pictures and sound, and ASSDA's H to increase the accessibility of historical documents as research tools, ASSDA could offer consultancy, advisory and curation services to the Humanities.

#### **Outreach**

ASSDA regularly performs tactical outreach services including providing advice on research proposals, ethics applications, data preparation, licensing and legal obligations, and establishing institution based data archives. One-on-one time is essential for this process to explain standards and train staff for data preservation across the project lifecycle. The service could be made more available and systematic, by having dedicated outreach staff to make contact with researchers in receipt of funding, provide lectures to research students, librarians and conduct nation-wide site visits to assist in local archive development.

Using the existing ASSDA skill base, other tools could be developed to assist researchers manage research data across the lifecycle. These tools might include a web interface using the DDI 3 metadata standard to collect and export metadata from the research proposal through to the archiving process to support the easy creation of documentation and repurposing of data files. Training manuals would be prepared, discussing data preservation obligations and strategies across the lifecycle and distributed to those in receipt of funding to collect data. Finally, the establishment of a professional association for data archivists in Australia – with links to relevant international bodies such as IFDO (the International Federation of Data Organisations – would be another possible enduring legacy of NCRIS funding.

**National Collaborative Research Infrastructure Strategy  
Platforms for Collaboration  
Australian National Data Service**

**Capacity Assessment Exercise**

**Introduction**

The Bureau operates as a single integrated national research and service organisation, serving the needs and meeting the responsibilities of Australia.

Data underpins much of what the Bureau does. It is the foundation from which the Bureau is able to provide its wide variety of services. The Bureau has some 57 manned stations in Australia, offshore islands and Antarctica. 50 of these stations provide upper air information through launching high altitude balloons between 1 and 4 times daily. The Bureau has over 500 automatic weather stations reporting from every minute to half-hourly to hourly, and another 450 co-operative observers who report every 3 hours to daily. There are over 7000 rainfall stations and over 130 drifting marine buoys. Our flood warning network includes some 700 river height sites and we have around 90 voluntary observing ships in Australian and adjacent waters. Approximately 40 aircraft have meteorological sensing equipment, which is automatically transmitted to the Bureau. In addition to the above the Bureau receives and processes satellite data from geostationary and polar orbiting satellites.

In total the Bureau generates approximately 1 terabyte of new data each day from observational data, satellite data, radar data and numerical weather prediction models.

The Bureau deploys a vast array of ICTs in the course of its operations, and faces the constant challenge to be a leading user of ICT and a manager and provider of data. The Bureau has developed extensive skills in real time data management and in service provision. For example, the Bureau's website accounts for 60% of the entire Australian Government's web traffic volume and the public downloads over 150 terabytes of data annually.

The Bureau has formed close links, and works collaboratively, with other research agencies such as CSIRO, universities and Defence, as well as with international organizations, such as WMO. The Bureau is continuing its role in high performance computing with the upgrade of its supercomputer in 2008 and a 5-10 petabyte-scale large scale data storage systems (2007). The Bureau's future vision includes a state-of-the-art Climate and Earth System Simulator (ACCESS) and participation in a comprehensive environmental data network. In addition, the Bureau is accommodating the Ionospheric Prediction Service (Space-weather) and (subject to agreement) the Water Information initiative of the National Plan for Water Security. The formation of a Centre for Australian Weather and Climate Research (CAWCR) with CSIRO will create a unique scientific capability in the areas of weather, climate, ocean prediction and earth systems science.

Given its current research and operational roles, experience in data management and provision of real-time services, existing and planned extensive ICT infrastructure, and its future direction, the Bureau believes it can contribute positively to ANDS.

### *Federated Services*

The Bureau wishes to play an active and central role in the initial support services and future layers as they are added. It is interested in common data analysis and visualization services, generic data quality assurance services, notification and data curation services, and the common data submission and presentation services. The Bureau has extensive experience and capability in the development and application of data standards. The Bureau could be a major node for NCRIS, including for the registration, location and access services. The Bureau is already positioned to undertake some of these roles with its BlueNet and BlueLink projects. The Bureau of Meteorology is in the business of data services as part of its national obligations and is interested in partnerships and other forms of collaboration that could build on (and with) such infrastructure.

### *Outreach Services*

The Bureau could also play a role in outreach services. The Bureau's niche is in communication with other government and non-government data initiatives, especially with regard to meteorology, water, numerical model output, imagery and weather, climate, hydrological and oceanographic research. The Bureau also could provide advice on mass storage, data curation, quality control and archives.

### *Stewardship Services*

The Bureau currently supports the continuity of access to a significant collection of data, which it makes available to the public, to other government and non-government organizations, as well as internationally. The Bureau is well placed to scale this service as required, particularly in relation to broader research access and dynamic infrastructure.



# NCRIS Australian National Data Services (“ANDS”) – proposed CSIRO Involvement

6th June 2007

## Purpose

This paper responds to DEST's email of 30<sup>th</sup> May 2007 requesting those who attended the ANDS workshop on the 29<sup>th</sup> May "to submit a statement of capability/interest in providing, contributing to or using services of ANDS"

### *At a glance.....*

CSIRO is willing to collaborate in/lead a consortium to create ANDS  
CSIRO is a national organisation & is well placed to deliver ANDS Outreach Services  
IM&T is a service oriented organisation (ITIL), capable of "hardening" the ANDS service delivery  
IM&T has a Project Management culture, required to successfully deliver ANDS  
CSIRO is a huge data source and consumer, and is committed to enabling Collaborative Australian Research across its divisions and close partners via its "One CSIRO" policy, its Flagship programs and its eSIM strategy

## ***Our Understanding***

A capability acquisition project for ANDS will draw on the successes, expertise and experiences of the MAPS, ARROW, DART, APSR, and AAF projects. These successful technology inputs will be a critical input to the total project, both initially and as they change/develop over time.

**The project will need to integrate these islands of capability into a "hardened" national service that will survive any initial hiccups to become a widely-used and trusted service for Collaborative Australian Research.**

## ***Our Approach***

The initial ANDS acquisition project (the "build") may not require the "central entity" to pre-exist. In fact the project may play a major part in helping determine the requirements for this central entity, which must be able to sustain the ANDS services. We believe that selecting a "central entity" right now may deflect energies away from the project's success.

CSIRO has operations across the nation. CSIRO comprises separate Divisions who increasingly need to collaborate with each other in order to solve the "Complex Science" problems can only be addressed by a multi-disciplinary approach. Consequently, CSIRO has already embarked on a number of internal programs (eSIM and Foundation) which are NCRIS-like in their objectives – fostering, enabling, supporting collaborative research across communities which are dispersed from both a geographic and a scientific discipline sense. A key enabler for CSIRO has been the adoption of the ITIL framework to ensure a consistent approach to service management and delivery, so that services provided are accessible, affordable and reliable. Early acceptance of ANDS by the research community is critical. Our approach would require a service-oriented outcome from the acquisition project (and introduction of the capability) to underpin this success.

CSIRO IM&T has a mature project management culture, and is well placed to take a leading role in the acquisition project, supported in consortium by the key contributing organisations responsible for AAF, MAPS, ARROW, DART and APSR. We believe the acquisition project must be driven as a formal project because it is too large, complex and important to be approached in a less controlled manner.

ANDS, or an ANDS like service, must succeed for CSIRO – must meet CSIRO's requirements for such a service – in order that CSIRO can realise for Australia the benefits of collaborative research

when dealing with the big problems and with its Flagship programs. This is why CSIRO wishes to play a major role in its acquisition.

### ***Our Proposal***

We propose a round table approach to solving integration problems between the various sub-capabilities in a collaborative way instead of the directive approach that would naturally result from a single management entity. Natural partners in the consortium are emerging, including Macquarie University, Monash University, Australian National University (APSR) and University of Queensland.

We propose a project team reporting through a project board to AERIC and lead by a CSIRO IM&T project manager, supported by a steering committee/project assurance team drawn from the consortium members. A key deliverable of the project would be to recommend an approach to the “central entity”, to manage the operations of the resultant ANDS capability.

## **Geoscience Australia Capability Statement Australian National Data Service (ANDS)**

Geoscience Australia (GA) is a potential contributor to the ANDS project to further enable the sharing of research data between relevant institutions and government agencies that make up the research community within Australia. GA has the capability that will benefit the ANDS project in multiple areas, these are described below:-

### *Federated Services*

GA's capability in this group of services is through provision of the discipline specific discovery services. The agency has managed and maintained geospatial discovery systems for the last 10 years, examples include the ANZLIC Australian Spatial Data Directory. The agency is in the process of implementing new discovery system that is ISO 19115 compliant. ANDS could benefit from the use of such discovery services as well as the knowledge gained during its creation.

### *Outreach Services*

GA's primary capability for ANDS is through the support of institutions in the implementation and creation of standard based data sharing infrastructure. GA's primary area of expertise is in ISO/OGC standards. GA is a technical member of several ISO/OGC standards committees and has experience in the formation and adoption of standards, particularly geospatial standards, within global communities.

GA recently formalised an organisational approach to data management. Institutions that wish to contribute to the ANDS may benefit from the aims, outcomes and standards implied in this approach. GA could provide support to the research community in use of such a methodology.

### *Contribution of datasets*

GA is the custodian for many of the national geoscience datasets. GA's core datasets can be freely downloaded or purchased for cost of transfer, however there are some limitations that ANDS should be aware of:-

- Current bandwidth is limited for the transfer of large data quantities (approximately 48Mb/s)
- Authentication and authorisation mechanisms established are consistent with the ACSI33 government guidelines
- GA's data holdings are stored on a mixture of online and offline media

With these limitations in mind, it is envisaged that our organisation could still make a substantial contribution to the content of the ANDS.

### *ANDS Usage*

The resource that would result from the formation of the ANDS would be of significant interest to GA staff. Many of these relationships and collaborations occur without this facility. However the formation of ANDS would introduce a level of efficiency whilst extending the opportunity for Geoscience Australia to further engage with Australia's research community.

## **Statement of Capability from iVEC**

iVEC, 'The hub of advanced computing in Western Australia' is an unincorporated joint venture between Central TAFE, CSIRO, Curtin University of Technology, Murdoch University and The University of Western Australia and is charged with encouraging and energising research in and the uptake of high performance computing, visualisation and large scale data storage by researchers in Western Australia.

In 2006, iVEC secured its third tranche of funding from the State Government, securing \$1.95 million per annum until 2010. A major component of this funding was to ensure that iVEC purchased a petabyte scale data storage system and make it available to researchers via a range of technologies including the grid. We are currently finalising the procurement process.

Thus the first capability that iVEC can bring to ANDS is access to a large data storage system (although the tender is not yet finalised, the system will have at least 400 TB of tape on delivery and will be expandable to at least 1 PB solely through the purchase of additional tapes).

iVEC will be actively involved in putting in place mechanisms for both the moving of data between remote sites as well as federating data across remote sites. As such, we would be very willing to get involved in any national federation services.

The final area where iVEC believes it can make a substantial contribution to the mission of ANDS is in the area of outreach. iVEC has been tasked by the State Government to provide outreach to researchers in Western Australia in the areas of high performance computing, visualisation and large scale data storage and the e-Research services that both bind these together and make them more accessible. As well as providing these services to the iVEC Partners listed above, we also have an industry and government uptake program that is actively engaging with companies and government agencies and could provide services through ANDS to this wider audience. Thus iVEC believes it is well placed to provide a home in Western Australia for outreach personnel to be directed through ANDS processes.

## DEST Funded Standards Capability – NCRIS ANDS Statement

Nigel Ward, Kerry Blinco – June 2006

DEST has developed a standards capability with a proven track record helping communities to solve real interoperability problems. The core of the standards capability is a team known as **Link Affiliates** that works with Australian communities, infrastructure projects, and with national and international standards organisations. Link Affiliates has been focused on education and repositories infrastructure, but in the past year has expanded to include e-research expertise. The team is distributed and portable. Organisationally, it is currently based at the University of Southern Queensland and has a close relationship with e-Research activities at the University of Melbourne.

As a result of this capability, Australia has an international reputation for influencing, adapting and adopting standards to build interoperable systems. The team participate in both formal ISO style standards organizations and more agile standards development consortia. The team take a broad view of the term “standard”, using a definition of “authoritative advice created through a consensus building and endorsement process”.

### Link Affiliates capabilities relevant to ANDS

#### Federated services

Link Affiliates can assist in the development of federated services. The team has experience working with existing SII projects including ARROW, APSR, DART, ARCHER, RUBRIC, MAMs, RAMP, AAF, and in piloting standards in demonstration projects including COLIS, PILIN, and FRED to establish best practice, policy, and governance for using the standards in ongoing service provision.

Link Affiliates activities of particular relevance to ANDS:

- PILIN - management and technical leadership to the PILIN persistent identifiers project.
- FRED project – development of metadata registry and discovery tools (including OAI harvesting tools).
- Standards activities relevant to the access federation (e.g. XACML, eduPerson).
- Standards development activities in metadata, ontologies, registries, discovery, repositories, content formats etc.

#### Outreach services

Link Affiliates can provide outreach services to the planning and early development phases of projects to:

- Identify, adopt and adapt (profile) standards to meet interoperability requirements.
- Engage with bodies developing e-Research standards.
- Provide advice on short-term and long-term planning for standards based interoperability.
- Use the e-Framework for Research & Education and other processes and tools for engaging communities with standards and strategic planning, create networks of expertise and document recommendations against data management issues.
- Producing best practice documents relating to interoperability issues and persistent identifiers

#### Stewardship services

Link Affiliates manages Australia's contribution to the international e-Framework for Research & Education. The e-Framework provides a methodology for creating, publishing and disseminating information about interoperability solutions, standards, their adoption and adaptation by local communities, and provides a knowledge base to record decisions, outcomes and best practice. The interoperability outcomes of many DEST funded infrastructure projects, including PILIN, FRED and APSR are being contributed to the e-Framework.

Link Affiliates manages and supports DEST's international partnerships targeted at the standards-based interoperability agenda. Partners include the JISC in the UK and the US Department of Defence. NCRIS capabilities could use these partnerships to access international experience and advice on building e-Research systems.

## **Macquarie University – ANDS Capability Statement**

Prepared by Professor James Dalziel

Director, Macquarie E-Learning Centre Of Excellence (MELCOE)

Macquarie University, through MELCOE, has been a significant contributor to the development of national systemic infrastructure for higher education and research. The following projects are relevant to ANDS:

### **MAMS (Meta Access Management System) - \$4.2M, MELCOE lead institution, 9 university partners, National Library of Australia, education.au & Telstra**

The MAMS project has developed and implemented infrastructure for federated identity and access management in the Australian higher education and research sector. Achievements include: the launch of a testbed Shibboleth-based federation in December 2005, which currently has over 900,000 identities from 13 universities and related organisations (eg, AARNET, ac3, VERSI); development of a Virtual Organisation management system which incorporates content storage for research groups (IAMSuite); development of a Shibboleth enabled Fedora repository (continued in RAMP project – see below); development of national shared services infrastructure such as Federation Manager and Shibboleth enabled MyProxy, and a Shibboleth WAYF (Where Are You From) service running in high availability mode with redundancy and backup.

### **ASK-OSS (Australian Service for Knowledge of Open Source Software) - \$200K, MELCOE lead institution, partnership with industry bodies (OSIA, Open Source Law and international partners – Oxford University “OSS Watch” JISC service**

ASK-OSS is an open source advisory service to the Australian higher education and research sector. It acts as a clearinghouse and also provides advice on open source software in the areas of development, selection, governance, business models, and legal/licensing issues.

### **RAMP (Research Activityflow and Middleware Priorities) \$2.9M, MELCOE lead institution, 4 university partners, range of international experts**

The RAMP project, DRAMA strand, has developed a Shibboleth enabled flexible web-based front end to the Fedora repository (“mura”) together with a generic XACML module for flexible authorisation for repositories and other eResearch systems. The RAMS strand is investigating and developing human collaborative workflow systems applicable to eResearch collaboration, repository submission and other human workflows, by building on and extending the existing architecture of LAMS.

### **Australian Access Federation (AAF), \$4.8M, UQ lead, Macquarie leads \$2.2M Shibboleth component**

The Australian Access Federation (AAF) is implementing a production ready Trust Federation for the higher education and research sector. The Shibboleth component of this work is being led by MELCOE, building on the work of MAMS.

### **Liaison with other Projects**

MELCOE has worked closely with a wide range of national infrastructure projects, including APSR, ARROW/DART/ARCHER, RUBRIC, Timesync (SIRCA) and others. It has also been involved in related work such as IMS Australia (standards), eFramework, PILIN, CORDRA and FRED. MELCOE has also developed close relationships with Macquarie-based NCRIS areas, such as APAF (5.1). MELCOE leads an international open source project (LAMS), including a community/repository website of 2200 members

### **Potential Capabilities for ANDS**

Macquarie University, through MELCOE, brings four key capabilities to ANDS:

- (1) Expertise in Federation Services for identity, access and authorisation, as well as related requirements for repository/metadata federation, persistent identifiers, distributed storage, etc
- (2) Expertise in repository system development and integration – as demonstrated by the development of mura, as well as the incorporation of secure content storage within IAMSuite
- (3) Expertise in leading, developing & implementing technical standards nationally & internationally
- (4) Leadership of large complex initiatives involving many partners across university, government and industry backgrounds – Professor James Dalziel has demonstrated the capability and passion required to lead and co-ordinate an initiative of the scale and complexity of ANDS.

# ANDS Capability and Interest Statement

## University of Melbourne – 4 June 2007

**Introduction:** The University of Melbourne, through activities directly related to the development of eResearch capability, is interested in participating in the Australian National Data Service [ANDS] both from the perspective of a prospective user and as an organisation that is keen to participate in its establishment and the development of its services. The University has established infrastructure including: a recently appointed Director of eResearch (a computer scientist); the eScholarship Research Centre [ESRC] (which works directly with researchers and project teams on issues relating to the sustainability and interconnection of research data and records); a close working relationship with the Victorian eResearch Strategic Initiative [VeRSI]; direct engagement with large scale data intensive international science collaborations (eg ATLAS); and an integrated informatic framework that captures and web-publishes information about research active staff at Melbourne ('Find an Expert'). Furthermore Melbourne is well positioned through Information Services, which brings together library, information and IT professionals, and is actively participating in the Australian Partnership for Sustainable Repositories [APSR] projects that are setting the foundations for ANDS. Melbourne, as one Australia's most research active universities, covers a broad range of disciplines from creative arts to high energy physics, producing research data and records of enduring value. It is committed to helping build a national framework of services that will enable this material to be responsibly managed and meaningfully utilised through time.

**Stewardship and Standards:** Melbourne, through the ESRC and its predecessors has unmatched experience over two decades in managing the transition of data and records (of all sorts) from the active research environment into the long term preservation environment. Managing complex access conditions and requirements has been necessary component of this stewardship service, as has training and consultancy services for a wide range of clients. Melbourne supports the systematic development and deployment of informatic standards as a key functional requirement for ANDS. Melbourne staff have been actively engaged in standards development at the international level in this field over the last decade, especially in archival standards for interoperability and interconnection between repositories (e.g. International Council on Archives), utilisation of library authority files; and recordkeeping metadata (ISO 15489, one staff member has a recently completed PhD investigating the 'clever use of recordkeeping metadata'). Recent activities include membership of the Australian Standards IT-019 committee and the development of a close relationship with the University of Southern Queensland PILIN project to extend the development of standards for persistent identifiers.

**Federated Services:** Melbourne, through VeRSI and the collaborative arrangements with Monash University (soon to include Latrobe University and the Victorian Department of Primary Industry), has technical capabilities in federated data storage and experience in the application of Shibboleth. The ESRC, continuing internationally-based work in the archives of science, and the global utilisation of national registers (ICSU – World History of Science Online), offers capability and experience in national knowledge base creation and long-term management. Melbourne has worked with the National Library of Australia and the National Archives of Australia on the 'context' approach to managing information in networked environments and is continuing this work under the 'People Australia' project. Furthermore Melbourne has worked with both open source and proprietary digital data repository systems and has migrated data between systems. What Melbourne brings to this area that is special is direct experience in building open frameworks that create multiple pathways for discovery and access but more importantly provide the pathways to the contextual information that is necessary to understand any component part, especially data sets.

**Outreach Services:** Melbourne has been working with APSR, VeRSI, overseas universities, the cultural sector, and active research projects to ascertain the range of outreach service that are required in the contemporary volatile information environment. These services include: developing data portals for life sciences; developing interfaces with instrumentation; assisting with data analysis, and the analysis of storage requirements; collaborations with National Library (People Australia); managing the knowledge in the nuclear energy and research area; working with many Australian universities and Imperial College London on archival management programs; experience in running an open access repository; and the with APSR being an active participant in the 'Online Research Collections Australia (ORCA) support network. This experience ranges from strategic policy advice through to specific handling of data in a broad range of settings.

**Governance:** Melbourne has shown significant leadership in this area and the Vice Chancellor keenly supports our continuing participation. Given our extensive experience, Melbourne is prepared to be actively engaged in the ongoing ANDS process, including the hosting of the 'head-office'.



**NATIONAL COLLABORATIVE RESEARCH INFRASTRUCTURE STRATEGY  
PLATFORMS FOR COLLABORATION  
AUSTRALIAN NATIONAL DATA SERVICE (ANDS)**

**MONASH UNIVERSITY CAPABILITY STATEMENT**

8 June 2007

## Capability statement

Monash University has a high level of capability in the areas of data management relevant to ANDS. This capability is based on a number of historic and institutional factors:

- Commitment to e-Research at the highest level of the university, including the Deputy Vice Chancellor (Research), the Director, Information Technology Services and the University Librarian, as well as recognition that this is a key to the University's strategic direction. This is evidenced by:
  - the establishment of the first e-Research Centre in an Australian university (2005);
  - the creation and implementation of an Information Management Strategy (2005) and subsequent selection of information management as one of only five key priorities in the University's Operating Plan (2006);
  - the creation of a Data Management Policy and Plan, endorsed by the e-Research Advisory Committee (2007) and now being trialled with research groups;
  - the establishment of large scale data storage facilities, a practical operating model and a sustainable funding model; and
  - the establishment of high performance computing capability and data storage facilities supported by national and state based services targeting institutional engagement with the Australian Synchrotron.
- Critical mass of skilled people, information management specialists (particularly Dr Andrew Treloar), e-Research support specialists, and computer scientists (particularly Professor David Abramson) leading to:
  - a demonstrated ability to work together to understand e-research and data management requirements;
  - a history of involvement with researchers and work on understanding their needs;
  - a highly skilled, capable, and multidisciplinary workforce underpinned by sound management practices to deliver original, challenging and innovative outcomes; and
  - a track record in research, development, and deployment of Grid middleware.
- Established record of involvement with state and national e-research and data management agendas:
  - three years' leadership of the national projects ARROW, DART and ARCHER;
  - leadership in the VERN initiative;
  - leadership in CAUL and CAUDIT, including conducting e-research and data management fora;
  - strong involvement with the VerSI and VPAC initiatives;
  - membership of relevant committees, including the e-Research Coordinating Committee and the Platforms for Collaboration Reference Group; and
  - the establishment of a data grid involving Monash, ANU and the University of Queensland providing large scale data storage facilities for researchers.

Monash believes that it can lead and respond to the following functional capability areas listed in the May 2007 ANDS Discussion Paper. This can obviously be elaborated on as required.

- Policy and planning
- Collaboration governance (ARROW, DART, ARCHER, VERNet, VerSI, VPAC)
- Project management
- Community development, outreach and support (ARROW Community)
- Repository leadership and management (ARROW, DART, ARCHER)

- Conceptualising a national discovery service (ARROW Discovery Service, established by the National Library of Australia)
- Metadata management (ARROW, and the recently formed Metadata Advisory Committee for Australian Repositories (MACAR))
- Data architecture (ARROW, DART, ARCHER)
- Data re-positioning (DART, ARCHER)
- Data curation continuum (ARCHER, ARROW)
- Addressing researcher needs (DART, ARCHER)
- Data management policy and planning (Monash Data Management Policy)
- Data storage provisioning and service management (Monash University Large Research Data Store)
- Storage management (DART, ARCHER).

### **Proposed involvement**

Monash University is interested in playing a leading role in furthering the ANDS goals over the next 12 months, in collaboration with other institutions. Based on its leadership of ARROW, DART and ARCHER, and its own institutional activities, Monash has a particular interest in leading the Stewardship (and associated outreach) components of ANDS. Monash acknowledges CSIRO's offer to lead a collaborative approach to ANDS and recognises the likely benefits of having a national institution in this role. Monash already has a strong relationship with CSIRO and is developing this around jointly operated data centres, which could logically extend to data management and high performance computing. In addition to other universities with key expertise, Monash sees the National Library of Australia making a strong contribution in the areas of resource discovery and preservation, and possibly persistent identifier service provision.

Professor Edwina Cornish  
Deputy Vice Chancellor

8 June 2007

## **NAA Capabilities for NCRIS Platforms for Collaboration - Australian National Data Service Activity**

### **National Archives of Australia (NAA)**

The National Archives of Australia (NAA) is the Australian Government's records and archives authority, and has extensive and well-respected experience in management of records (data) in many forms, both analogue and digital. NAA is an enabler of good records management in government agencies, providing standards and guidelines to assist records creators and to ensure that agency staff create complete and reliable records that remain accessible and useable over time.

NAA understands and promotes the view that it is vitally important for records/data creators to have in place proper strategies and policies that ensure the integrity, durability, security, useability, meaning and authenticity of digital data whatever its form. To this end NAA has published a wide range of guidelines and better practice standards that cover topics such as managing business information (the DIRKS Manual), recordkeeping metadata schema, discovery and access metadata, ontologies, records appraisal and disposal, and policies and strategies for archiving of web-based resources. More recently NAA has issued guidance on creating, managing and preserving digital records, and on information authentication issues.

NAA has developed and implemented an operational digital archive that stores, manages and preserves digital objects created by government agencies. The features of the digital archive are: use of open source technologies; conversion of objects to a standard range of archival data formats to produce authentic, reliable, and durable digital objects; and a suite of NAA-developed tools that manage and undertake the digital preservation processes.

NAA has unique expertise in issues of data management and sustainability and is interested in working collaboratively with data management initiatives and other information professionals who will contribute to the ANDS program.

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

NAA is interested in contributing to a number of the ANDS service areas. In particular NAA wishes to contribute to the following ANDS service areas:

- online collections registry and discovery service –NAA has been contributing to development of a National Online Archival Network for discovery and access to archival resources across Australia;
- metadata and ontology scheme registry – NAA has long involvement in the development of ontologies such as AGIFT, and metadata schema, in particular the AGLS Metadata Standard for resource discovery AS5044. In the context of our leadership role in relation to the AGLS Metadata Standard NAA has participated actively in the metadata working group of the National Data Network;
- Outreach services – as the Commonwealth's archives and records management authority NAA has extensive experience in information management issues and practice. For example, NAA is leading the Australasian Digital Recordkeeping Initiative collaborative activity to harmonise strategies for managing digital objects across Australia and New Zealand. NAA would make an essential and important contribution to all components of this service area both strategic and tactical;
- curation and collection management services – NAA has developed and implemented a digital archive for storing, managing and preserving over time digital objects created by government agencies. NAA can offer a number of tools specifically developed to assist and manage digital preservation processes, including the XENA preservation tool that has generated wide-spread interest and some uptake outside the Commonwealth.

## **Capability Statement for ANDS (NCRIS PfC)**

**on behalf of the PACs: QCIF, ac3, VPAC, TPAC, SAPAC, iVEC; ANU and CSIRO**

### PACs role in ANDS

The PACs (Partners for Advanced Computing) are the original state-based APAC Partners that, along with two other national organizations, ANU and CSIRO, created the first national, collaborative advanced computing program in 2000.

The PACs are well aligned with the NCRIS Guidelines, having a service focus – in particular, the PACs support, rather than do, research; by leveraging modern ICT. Providing service for researchers is the core mission of the state-based PACs. Together they include more than 25 Australian universities in their membership and have attracted continuing endorsement and financial support from 6 state governments. They have demonstrated their cooperative national role in the ICI Component of PfC.

The PACs are unarguably the leading national federation of advanced computing research support and are well down the path of extending that support to research data, in response to community demand. The PACs have a proven track record and are logical national “service providers” to both ICI and ANDS. Universities, in the main, are not set up to provide support to others, providing a clear distinction between the state-based PACs and the research institutes and universities they serve. We acknowledge that there is additional, specialist expertise required in developing the full ANDS infrastructure (eg. related to registries, curation, metadata, ontology, etc) that may lie beyond the PACs.

PACs Capabilities: Each PAC has submitted individual Capability Statement for ANDS (attached). These list their substantial data archival facilities and multiple data-related R&D activities and services. This combined archival capacity already exceeds 4,000 TeraBytes (TB), with > 500 TB under active management; and represents a substantial fraction of the national research data storage capacity. To date funding has been provided by state governments, member universities, ARC, DEST SII, etc. Several of the PACs are working in a coordinated manner to provide upgrades of ~ 100 TB each per annum, while the remainder are installing substantial new capacity in the coming year, to keep pace with emergent demand for storage of scientific data. Recent major grants have led to more coordinated activities, systems integration and data interchange. While the original PACs’ focus has been on large scale scientific data (from Astronomy, bio-informatics, geo-sciences, oceanography, etc), at least two sites (ANU, UQ) have a significant effort devoted to social sciences data; while ac3 has a long association with financial data (via SIRCA); and VPAC has a strong program in Health Informatics (MMIM/ACG). The PACs now are increasingly supporting the humanities and social sciences, in response to opportunities and demand.

### PACs bring to ANDS:

- ❖ Operational archival storage and hosting capabilities, suitable for the ANDS constituency;
- ❖ Practical experience with migrating scientific data across evolution of the underlying storage technologies as they change over time (thereby avoiding “digital rot”);
- ❖ capabilities in serving data through the various services (eg Web services, OpendDAP, content management systems)
- ❖ Skills in portals, workflows and data-visualisation pipelines; and portals that allow access to data services, some with knowledge discovery tools (crawlers, OAI harvesters);
- ❖ Skills in data federation (eg via SRB, OpenDAP), replication and data movement;
- ❖ Experience in bringing modern data methodologies to new communities (humanities and social sciences);
- ❖ Outreach operations: originally focused on HPC, now evolving to support scientific data, particularly across the software stack from underlying hardware to end users
- ❖ Skills in creating, managing and working with a national-scope virtual organisation.
- ❖ Software project management skills
- ❖ Experience in data migration and validation, and creation of data and meta-data schemas;

9 March 2007  
Ref. NLA 07/508



Dr. Michael Sargent, AM  
Chair, NCRIS Committee  
Department of Education, Science and Training  
PO Box 9880, Canberra City  
ACT 2601

CANBERRA ACT 2600  
AUSTRALIA  
TEL +61 2 6262 1111  
FAX +61 2 6257 1703  
TTY 1800 026 372  
[www.nla.gov.au](http://www.nla.gov.au)  
ABN 28 346 858 075

Dear Mike

**RE: INVESTMENT PLAN - PLATFORMS FOR COLLABORATION**

I recently met with Rhys Francis to discuss the draft investment plan for the NCRIS Platforms for Collaboration capability.

The National Library is encouraged by the shape which this investment plan is taking. We welcome the proposals to establish ongoing governance structures such as a National Research Infrastructure Coordinating Committee, an e-Research Infrastructure Program Office and an e-Research Forum. We support the appointment of an Executive Director and permanent staff to carry forward the investment plan.

The Library supports the establishment of the proposed Australian National Data Service (ANDS). The Library was fortunate to have one of its executive staff (Dr Warwick Cathro) as a member of the recent PMSEIC Working Group on Data for Science. It appears that ANDS will be an important driver for achieving the recommendations made to PMSEIC, and for ensuring that Australia's research data is well managed as a national asset.

The National Library has been an active participant in some of the projects funded by the Systemic Infrastructure Initiative, such as ARROW and APSR. We see the Platforms for Collaboration capability as a mechanism for ensuring that Australia builds on the valuable work undertaken by such projects.

The Library welcomes the attention given in the draft investment plan to issues such as the stewardship of research data, persistent identifier services, provision of advice to researchers on data management practices, and improved registration and discovery of data sets.

With respect to the future operation of this research infrastructure capability, the National Library would be willing to participate in the NCRIS or ANDS advisory processes if this is deemed appropriate by the NCRIS Committee. The Library also believes that it has the experience and expertise to make a contribution to some of the proposed infrastructure services. In particular, the Library could make a contribution to:

- registration and discovery of data sets, building on the Library's track record in establishing other national resource discovery services;



- the sustainability and data curation services of the proposed ANDS, building on the Library's longstanding work in digital preservation, web archiving, and our contribution to the sustainability aspects of the APSR Project;
- development of standards supporting inter-operability of data repositories; and
- A national Persistent Identifier infrastructure service, building on the Library's work with its own persistent identifier service and its encouragement of good practices to manage web resources for persistent access.

The Library is also interested in the development of collaborative tools that support research in the humanities. To take an example, we are interested in the development of generic tools that can support online publishing with scholarly annotation and linking to related resources and source materials such as those held in the collections of libraries and other institutions.

Yours sincerely

Jan Fullerton  
Director-General

cc Dr. Evan Arthur, Group Manager, Innovation and Research Systems Group, DEST.  
cc Dr. Rhys Francis, NCRIS Platforms for Collaboration Facilitator.

## QCIF capability statement for ANDS (NCRIS Pfc)



QCIF Ltd is a consortium that provides advanced computing, cyber-infrastructure and related e-research services in Queensland. Its founding members are six of the Queensland universities (UQ, QUT, Griffith, JCU, CQU and USQ). In 2006 the organisation changed its name from the long-used QPSF, founded in 2000. The mission of QCIF is to deploy and support advanced computing infrastructure and e-research services to support Queensland's Researchers and Industry. QCIF is one of the so-called PACs and has been actively involved in APAC and now Pfc, especially the ICI Component.

### Data related R&D and Services:

Member universities host significant R&D projects relevant to data: UQ and JCU are major partners in DART and ARROW; QUT hosts the e-Law project and has led the APAC Portals project; UQ is a partner in APSR; UQ and JCU are active in tele-instrumentation (CIMA, etc) for materials characterisation. A number of these projects likely will be detailed in separate submissions to DEST. QCIF has staff dedicated to supporting data users in a "data grid" environment: for Storage Resource Broker (SRB) services; for data-visualisation, user interfaces and workflows in support of data analysis (at UQ, QUT, JCU and Griffith) and for new non-expert users (eg. archaeology and geography at UQ). Additionally QCIF has assisted AIMS in establishing Sensor Networks for monitoring the Great Barrier Reef. QCIF provides data support to IMOS (via AIMS & JCU), to Bioinformatics (ACB & QFAB) and to AusScope (via the former ACcESS MNRF); Griffith hosts the local mirror of the Ensembl bio-informatics database. QCIF hosts a large number Access Grid (AG) nodes for collaborative services, supports their operation nationally, and leads Australian AG development effort – this includes native sharing of data within AG sessions via SRB.

### QCIF HPC and Data Resources:

QCIF's supercomputers are distributed across 4 university sites, comprising ~1,300 processors (out of ~5,000 nationally in PACs) with an aggregate peak speed of ~ 4 Tflops. QCIF hosts, at UQ, one the major research data archival systems nationally – the others being at ANU, Monash and CSIRO/BoM (with iVEC and SAPAC also installing systems in the near future). At UQ the StorageTek PowderHorn Tape archive currently holds 220TB of scientific data (with 1,200TB capacity), along with 15 TB disc, managed via SGI's DMF Hierarchical Storage Management software. A regional facility at JCU is a 95 TB StorageTek L180 Silo, also using DMF. These systems have been established with substantial university, QCIF and ARC LIEF funding. Currently QCIF is upgrading this data capacity at 100TB pa (tapes, matched ~10% disc), to keep pace with demand. Major data holdings are for: bio-informatics, spectroscopy (MRI and presently crystallography), geo-sciences, marine sciences, satellite imaging, health sciences (MRI, immunology), with data services to AusVO and social sciences.



[www.qcif.edu.au](http://www.qcif.edu.au)

(B. Pailthorpe, CEO, June 8 2007)

Professor Brian Fitzgerald and Dr Anne Fitzgerald

The QUT Legal Research project proposes to contribute to the ANDS project by addressing the legal aspects of the management of research data which are relevant to each of the key elements of the ANDS structure:

- **Stewardship services:** Management of legal and licensing issues relating to research data is fundamental to the curation and management of digital data, particularly in the context of collaborative, publicly-funded research projects which seek to improve the mechanisms for data access among research groups. It is essential for researchers and data custodians (eg repositories) to have a practical understanding of a range of legal issues including ownership of intellectual property and other rights in data and databases, authorisation to use data assets owned by other parties, and restrictions applying to the distribution or use of data.
- **Federated Services:** An efficient federated data network demands a common understanding and uniformity of practice among participants with respect to the management of legal rights in data, so that practices are compatible not only among the various distributed research data centres in Australia but also with the practices of data centres in key overseas jurisdictions.
- **Outreach Services:** As the legal aspects of data management are complex and not well understood within the research sector, assistance is required to develop, through selected examples and case studies, effective strategies to deal with the legal aspects of data generation, access and use. Information about strategies for data management must be easily and freely accessible by researchers and data custodians throughout the federated network on a real-time basis. For this objective to be achieved, an appropriate range of web-based resources needs to be developed - including practical guides, toolkits, checklists and template documents – and ongoing legal and technical support is required to ensure that these information resources continue to be updated, useful and accessible for the duration of the ANDS project.

The QUT Legal Research project has completed a report, *Building the Infrastructure for Data Access and Reuse in Collaborative Research: An Analysis of the Legal Context of Research Data Access and Reuse* (June 2007). The report examines the broad legal framework within which research data is generated, managed and used and provides a basis on which to develop stewardship, federated and outreach services specifically targeted to the requirements of research centres and data repositories within the ANDS structure. As well as providing an overview of the operation of relevant laws (eg copyright, contract, privacy and confidentiality laws) in defining rights to access and use research outputs, the report describes and explains current practices and attitudes towards data sharing, and surveys international and national policies on data access and use. The report describes best practice strategies for research projects in organising, preserving and enabling access to and reuse of data.

The report's principal recommendations are of direct relevance to the management of research data within the ANDS structure:

- Further analysis of existing data access policies, principles and practices in Australia and overseas should be undertaken with a view to formulating template data access policies and principles for use in the Australian research sector.
- Data Management Plans (DMPs) should be developed to address how data is to be managed in accordance with relevant legal requirements. DMPs should address how research data and databases will be preserved and maintained in the long-term and how this will be funded, how data is to be made accessible to the public and who will be primarily responsible for the management of the database and data.
- Data Management Toolkits (DMTs) should be developed to assist researchers in collecting and dealing with data in accordance with the DMP, how to deposit data into the database and how to manage their data in accordance with regulatory frameworks applying in their area of research.



Prof. Anthony G. Williams, Director  
SAPAC  
Level 1, Physics Building  
University of Adelaide  
SA 5005, Australia  
*Telephone:* +61 8 83033546  
*Facsimile:* +61 8 83033551  
*E-mail:* Anthony.Williams@sapac.edu.au  
*Web:* <http://www.sapac.edu.au>

## SAPAC Statement of ANDS related capabilities

SAPAC is an unincorporated joint venture of the three South Australian universities, Flinders University, the University of Adelaide and the University of South Australia. Its primary mission is to provide services and infrastructure to enable 'Discovery, Innovation and Collaboration through e-Research'. SAPAC is funded by contributions from the three universities and the State Government. By the end of 2007 SAPAC will have been transformed and subsumed into e-Research SA (eRSA) and as part of this process the State Government will become one of the founding partners in addition to the three universities.

SAPAC manages more than 8 Teraflops of high-performance computers including its newest computer "Corvus", which is a 6 Teraflop quad-core SGI Altix XE1300. Corvus is a general purpose, high-performance computing platform and, as an Australian dedicated research computer, is second in compute power only to the APAC National Facility. In addition, SAPAC manages two visualization facilities, remote video-conferencing rooms, haptics services, massive RAID and tape silo data storage facilities, high bandwidth interconnect via SABRENet and the new dual 1 Gbit/sec SAPAC-VPAC research network link as well as offering consulting, expertise and training courses to the R&D community.

Of particular relevance to the ANDS capabilities is the experience and track record of SAPAC in e-Research services provision and its recently established South Australian Sustainable Repository (SASR). SASR is a managed, distributed, mass research data storage facility primarily intended to service the South Australian research and development (R&D) community. Although presenting as a single, unified system, SASR has been implemented as three physical nodes located at the three South Australian Universities. These operate as gateways for researchers from their respective Universities to access transparently SAPAC's broad range of e-Research resources and, specifically, SAPAC-managed research data storage and related services. SASR will over time be expanded to include other R&D groups, from government and industry, by adding nodes and additional data storage capacity. SASR complements and augments existing and planned individual institutional and discipline-based digital repositories and is an exciting new service for the South Australian research community. At present SASR consists of more than 40 Terabytes of RAID disk and a 308 Terabyte tape silo currently being installed and fully integrated into SAPAC's information infrastructure and HPC fabric using SGI's DMF file-management system.

SAPAC is funded in part by the ICI component of NCRIS PfC specifically to manage data storage and movement and will be hosting several large datasets including, for example, bioinformatics data (in particular Plant Functional Genomics and Plant Phenomics), drama performance data (AusStage) and humanities datasets, as well as nationally significant health informatics data. Thus SAPAC/eRSA is the natural partner to provide both federation and outreach services for ANDS in South Australia as well as to provide stewardship for some nationally significant data sets.

## SIRCA Capability Statement for ANDS

SIRCA has been specifically structured as a multi-university cooperative to maximise opportunities for collaborative activities directed at building and enhancing e-research services for the financial, economic and broader social sciences academic community. Supported by DEST (via SII) and in collaboration with its 29 Australian and New Zealand Universities along with many government and industry partners and a growing number of international universities, SIRCA has emerged as a global leader specialising in financial e-research services. SIRCA is well positioned to leverage its substantial and proven infrastructure for the benefit of the wider e-research community.

SIRCA has proven capability in the following areas:

- SIRCA is the **trusted custodian** for major global and local protected data sets comprising fine grain (mili-second) historical time series financial, economic, environmental, and social events. Most of this event stream data is sourced from real-time market transaction and news feeds and is probably the world's largest social science related data repository in the world (capturing between 20,000 and 100,000 messages per second).
- Significant experience in **liaising with global commercial data providers** on behalf of academics including commercialization, legal and intellectual property issues. Effectively managing relationships with data providers on behalf of academics has been a key to SIRCA's success.
- Proven expertise in the **management, access and delivery of historical time series event data** including expertise in the coordinated analysis and curation of multiple data types, such as combinations of numerical, textual, geographical and time-series data.
- An "industry strength" **application development capability** comprising around 20 full-time software engineers and quality assurance professionals working in a best practice **software development, test and production environment**. This web services model also incorporate agile development activities (based around Ruby on Rails technology) and is supported by on-line and face-to-face end-user support and training. Leveraging this capability, a new innovation centre is being established to facilitate rapid technology transfer and knowledge sharing among collaborating institutions.
- **Secure and resilient commercial grade production environment**. During a recent global audit by KPMG on behalf of Reuters, SIRCA received one of the highest security and data management ratings (even compared to internal Reuters infrastructure). AC3 is the current primary site with a second **disaster recover ("mirror")** site currently under development and due for completion in October. This will include a new Network Area Storage architecture to be implemented across both sites.
- A proven **governance and professional management** capability which is geared towards self-sustainability and enabling a tight feedback loop between end-users and in-house technologists.
- SIRCA has a well established and rapidly growing **cross-disciplinary end-user network** which is now global in scope. As such it has well developed capability in dealing with institutional subscription, technology and legal requirements in many jurisdictions.

SIRCA is well positioned to provide "satellite" services to ANDS that leverage and extend the above capabilities and thereby continue its strong participation within the broader e-research infrastructure community. This could range from provision of data management consultancy services through to shared data storage and web service application development for collections without a natural home or the appropriate scale to deliver a high standard of e-research service.

SIRCA is actively investigating new collaboration on a number of fronts – particularly in the broader social science domain and where there exist common data management issues or need for **commoditised** e-research solutions.

The University of Queensland (UQ) is willing to act as either a central ANDS entity or as a “shoulder” service provider. UQ is used to working collaboratively with others in the sector and we believe that ANDS will only be achieved if a series of federated service providers gain both the trust of their constituency (such as universities or research facilities in their region) as well as that of other service providers. We believe that one of the service providers could act as “primus inter pares” and thus be the central entity. UQ would be prepared to act in this capacity or to act as a service provider if that is considered a more appropriate contribution to the sector. ANDS must integrate very closely with the Australian Access Federation (AAF) and there may be some value in bringing the two entities together.

No one institution or organisation has a monopoly on the necessary skillsets or could hope to provide, on its own, everything necessary to deliver the ANDS Services, which is why ANDS will be a co-operative effort by a number of institutions. It would be surprising if that effort did not include ANU, CSIRO, Macquarie, Melbourne, Monash as well as UQ but it is likely that there will be a number of others providing specialist skills. Each will have their own strengths; UQ believes its strengths to include:

## Federated Services

UQ has significant expertise in the development, deployment and support of FEZ, a web based digital repository based on the open source Fedora software, as a major part of the APSR project. This has already been deployed internationally in 5 countries. UQ is a leader in ORCA, a sub-project of APSR, which is the coordinating support for online research collections and is also the lead university for the AAF project. UQ brings to the AAF, very significant expertise in access services, supplemented strongly by the expertise of AusCERT. UQ/AusCERT are also leading the auEduperson Working Group, which aims to develop a common identity schema for Australian universities. The university has a representative on the management committee of the National Library Resource Discovery Service and has specific experience in both OAI harvesting and repositioning data between repositories through building interfaces between FEZ and LCMSs and Research Output repositories. In addition, the UQ ITEE eResearch Group are internationally-acknowledged leaders in the fields of metadata schema design; schema registries; interoperability; data management and curation; federated search and retrieval; preservation services - with over 20 years experience in the development of data archival tools and services for major institutions and significant involvement in both international standards and initiatives ( including Dublin Core, MPEG, W3C, OAI-ORE, NSDL, CODATA, DELOS, UK DCC, JISC, UKOLN).

## Outreach Services

UQ has a substantial capability in Outreach activities. Examples include the APSR consultancy service that is provided to other universities, the training service provided by AusCERT and the UQconnect ISP service. AusCERT has a strong reputation and provides subscription based 24x7 internet security services to universities, governments and private industry; all Australian universities are members and have a trust relationship with the group. AusCERT undertakes very significant training activities with courses being delivered to universities, private industry, governments and police. The group has operated recently in 27 countries with one of the most recent engagements being the training of the national CERTs in Mexico, Chile and Peru, on behalf of APEC. As UQconnect, the university is a full Internet Service Provider (ISP) delivering services to students, staff, alumni and the general public. With over 65,000 users, UQconnect has considerable experience in providing service to users ranging from home internet users to users of High Performance Computing Systems. As an ISP, the university already runs a help desk service with extended hours and is planning to enhance this to a 24x7 capability within the next 18 months.

## Stewardship Services

UQ has considerable experience in the delivery of curation and collection management services and the dissemination of best practice, through its work on FEZ. As a data service provider, the university already manages over a petabyte of data storage, both for its own use and for the Queensland Cyber Infrastructure Foundation (QCIF). The majority of QCIF data storage is located at UQ and UQ ITS supports it under a formal Service Level Agreement (SLA). As the lead university in developing the proposal for a federated, collaborative and highly distributed data service involving all the universities in Queensland and Northern NSW, (QADDS), UQ is experienced in brokering solutions for data services.

## Governance

UQ ITS has adopted ITIL and is formally accredited under the ISO 9001:2000 international standard with full external audits every 6 months to maintain this accreditation. There is an established project office, based in ITS, which is accredited under these arrangements has adopted a formal project management methodology. We see the governance of ANDS being collaboratively achieved through the formation of a steering committee of major providers with the central entity providing a chair and overall project manager. The overall Project Manager would be supplemented by project managers for sub-projects undertaken by other providers. UQ is capable of providing significant project management expertise to support this activity and believes that there may also be potential for a link with the governance process that UQ is developing for the AAF project.

**UQ contributes to a large number of collaborative projects within Higher Education and Research. These include:**

**Australian Access Federation (AAF), CAUDIT/Grangenet PKI (CPKI), eSecurity Framework (eSEC), Middleware Action Plan and Strategy (MAPS), auEduperson Working Group (auED), Australian Partnership for Sustainable Repositories (APSR), Australian Research Enabling environment (ARCHER), Data Acquisition accessibility and annotation eResearch Technologies (DART), Online Research Collections Australia (ORCA), Regional Universities Building Infrastructure Collaboratively (RUBRIC)**

**And contributes significantly to, or manages, the following collaborative groups, organisations or conferences:**

**Australian Computer Emergency Response Team (AusCERT), Queensland Regional Network Organisation (QRNO), International Systems Security Professional Certification Scheme (ISSPCS), Queensland Advanced Distributed Data Service (QADDS), Questnet, Queensland Cyber Infrastructure Foundation (QCIF), Queensland Facility for Advanced Bioinformatics (QFAB), eResearch Australasia Conference 2007**

Also Nick Tate is a director of both AARNet Pty Ltd and HES Pty Ltd, Graham Ingram is a member of the AUDA names panel, Professor Bernard Pailthorpe is CEO of QCIF and Professor Mark Ragan is the chair of the board of QFAB

**Bold type** indicates projects or activities where UQ is or has been the lead institution

---

**UQ has a number of experienced staff who are willing and able to contribute to the provision of ANDS. These include:**

Professor Paul Bailes [ARCHER] ,Andrew Bennett {APSR, ORCA}, Ron Chernich {DART, ARCHER}, Chris Christoff {QRNO}, Professor Jane Hunter {APSR, DART, ARCHER, RUBRIC, GrangeNet}, Graham Ingram {AUDA}, Robert Mead {AAF}, Dr Rodney McDuff {CPKI, eSEC, AAF, auED}, Patricia McMillan {AAF, MAPS, auED}, Mark McPherson {AusCERT}, Professor Bernard Pailthorpe {QCIF}, Viviani Paz {CPKI, eSEC, AAF, auED}, Professor Mark Ragan {DART, ARCHER, QFAB}, Nick Tate {CPKI, eSEC, AAF, MAPS, DART, ARCHER, AusCERT, ISSPCS, QRNO, QADDS, AARNet, HES}, Keith Webster {APSR, ORCA}, Professor Xiaofang Zhou {DART}, John Zornig {eSEC, AAF}

{ } indicates participation in projects or activities

## The University of Sydney

### Statement of capability / interest in participating in the Australian National Data Service (8 June 2007)

This statement outlines the capabilities and interests of the University of Sydney to participate in ANDS. The University is open to further discussion about these capabilities, and while this statement reflects no commitment on behalf of the University it does indicate potential areas of participation in ANDS.

#### ANDS central managing entity

The University has no interest in leading or hosting this entity as an individual institution. In many ways this central managing entity may function more effectively as a distributed entity. The University may be interested in being a partner in a collaborative entity and participate in outreach and federated services.

The University would be interested, in collaboration with a proposed NSW e-Research support centre, in forming or contributing to a state hub for co-ordinated NSW data service activities for ANDS.

#### General stewardship services

The requirements of stewardship in the ANDS document need more detail and definition. The University does have expertise and skills that can contribute towards the general stewardship role of ANDS. These include experience in managing large data sets in local and networked environments through a number of major research facilities, participation in APSR programs, such as the development of SUGAR (Sustainable Guidelines for Australian Repositories), participation in the ORCA project, and local leadership in digital project planning and standards adoption.

#### Stewardship of satellite services

The University of Sydney has proven capabilities to lead or play a major role in developing protocols and providing stewardship services in a number of discipline/community groupings, as well as capabilities to contribute to others. Groupings where Sydney may lead or play a major role include:

- Microscopy (Capability 5.3) – Sydney has significant expertise and reputation around microscopy and associated research activities, including nanostructures, through micro-imaging and analysis services. This expertise includes consultancy and training services as well as facilities. Capabilities are provided through the [Electron Microscope Unit](#) which incorporates the Australian Key Centre for Microscopy and Microanalysis, and the Nanostructural Analysis Network Organisation Major National Research Facility ([NANO-MNRF](#)). Participation would also be in collaboration with the University of Queensland through the NCRIS National Microscopy and Microanalysis Research Facility
- Culture, arts and humanities – Sydney has strength, expertise and reputation in creating, managing, archiving, curating and providing access to data across the humanities, including text, GIS mapping, images, audio and video. This is demonstrated in leadership through the [Australian e-Humanities Network](#), and consolidated services in supporting innovative research practice in digital cultures. Capabilities include [PARADISEC](#), [Archaeological Computing Lab](#), [SETIS digital library collections](#), image, audio video services in fine arts, performance, music and linguistics, publishing services and extensive collaborative networks in Australia and overseas.
- Clinical data (5.7) – Sydney hosts several health data services and has an interest in ensuring protocols for the stewardship and appropriate management of these datasets within a national context
- Plant sciences (5.2) – Sydney is active in developing institutional and collaborative services supporting the imaging, description and taxonomies, storage and publishing of botanical data. Local capabilities include the development of the [eBot data](#) services and innovative publication through [eFlora](#).
- Other areas of capability and interest include Bioinformatics and Genomics (5.1) through Sydney Bioinformatics (including the [Australian National Genomic Information Service](#)); Marine studies (5.12) through the [University of Sydney Institute of Marine Science](#); Astronomy (5.10) through [Sydney University Molonglo Sky Survey](#); and high-energy physics through participation in the [ATLAS Experiment](#) and through [AUSHEP](#)

#### Contact

Ross Coleman, Director Sydney eScholarship  
[r.coleman@library.usyd.edu.au](mailto:r.coleman@library.usyd.edu.au) ph 029351 3352

## **University of Tasmania CAPABILITY STATEMENT in relation to the ANDS (Australian National Data Services) Discussion Paper**

*The University of Tasmania hosts TPAC, BlueNet and eMII, each of which have a corpus of expertise both locally and in national working groups convened under their aegis. Both TPAC and BlueNet, of which the Australian Ocean Data Centre Joint Facility (AODCJF) is a partner, have a strong record of activity and delivery in data management and curation, data delivery mechanisms, data product development, and in the development of strategies to disseminate capacity in these areas.*

*Identifying and assisting researchers and their organizations to manage their data and resolve research data management issues has been a core focus of BlueNet, and is also a primary task for eMII. BlueNet and eMII contribute significantly to development of the technologies and practices supporting the Australian Ocean Data Network (AODN), which serves as a focus for all Australian marine science data, including all IMOS data.*

BlueNet project office staff and BlueNet community members on AODCJF working groups are closely involved in the review and extension of data and metadata standards, strategies for their development and adoption, strategies and solutions for interoperability, and discoverability structures and services. The focus on standards covers collection and storage of data and metadata, identification of ISO-compliant metadata standards which can also meet the needs of the marine science research domain, and standards-of-practice to optimise data searching, data handling, curation and delivery, including web-service delivery. BlueNet has developed a process to meet workflow needs, and deliver relevant IP and use-constraint information. It has developed initial metadata translator and brokerage services. TPAC has played a lead role in developing tools and plugins to enhance discovery of and access to earth systems sciences data collections. Meeting these needs of the marine science community has required definition and development of appropriate architecture while also facilitating interoperability with related disciplines.

BlueNet and TPAC have also developed key technologies extensible to other domains. For example, TPAC has built a digital library offering content-stream access to self-describing file formats (e.g. NetCDF) of data from a range of sources, and BlueNet has led the development of the Metadata Entry and Search Tool (MEST), which is a core component of the AODN infrastructure. The MEST offers the capacity to create standards-based catalogue records, manages user-levels, provides direct links to data and related documentation and publications, web-mapping services and focused search capacity including federated searching.

A key part of this activity has been to develop expertise within the community, and this includes development and provision of a range of support materials, including toolkits, and feedback processes. BlueNet staff have taken a lead role in developing and delivering training and related support to both use the technologies of the AODN and in relation to standards awareness and adoption. BlueNet has broad experience in outreach to individual researchers, research groups and data-provider organizations in promoting good data management practice, integrating data management activities, and advising and providing tools to assist in automation of metadata creation, extension, maintenance and re-use.

BlueNet and TPAC develop and supply services to the marine domain, as outlined above, and play a leadership role within part of a broader consortium for aspects such as federation services and standards (data and metadata/ontology), and interoperability development. It would be efficient use of resources for these roles to continue within the marine community, and to assist other domains in a similar manner.

## List of Acronyms

eMii	eMarine Information Infrastructure	Funded under the IMOS stream of NCRIS
IMOS	Integrated Marine Observing System	
BlueNET	Australian Marine Science Data Network – integrating the Higher Education Sector into the AODN	Now included in the AODN
TPAC	Tasmanian Partnership for Advanced Computing	APAC member
AODCJF	Australian Ocean Data Centre Joint Facility	Marine science data expertise and practice from 6 Commonwealth government agencies: Australian Antarctic Division, Australian Institute of Marine Science, CSIRO-Marine and Atmospheric Research, Bureau of Meteorology, Royal Australian Navy and Geoscience Australia
AODN	Australian Ocean Data Networks	

## VPAC Capability Statement – for ANDS

The following capability statement covers the capabilities that VPAC can provide to ANDS, in collaboration with its Member Universities and other organizations such as VeRSI. The capabilities that VPAC can provide to ANDS are classified in a hierarchy, from basic hardware for data storage, to storage management and applications support. The capabilities described below are based on current VPAC employees and projects at VPAC that are “data-centric”, as opposed to “compute-centric” (though any project always has a blend of data and compute components). VPAC employs approximately 50 technical staff.

VPAC is involved in several major data projects, especially in the domain of Health Informatics

- ❖ Health Informatics: The MMIM project and its successor the Australian Cancer Grid (VPAC is a major contractor on this project, providing services ranging from Project management to software developer)
- ❖ VICNISS: Federation of hospital data sets for reporting on hospital acquired infection (being deployed across Victoria and designed and developed by VPAC)

VPAC also has major projects with data components in Engineering and Geospatial sciences.

### Data Hardware Capabilities (Storage)

- ❖ VPAC has a large NFS storage system with a major acquisition of new storage in progress (~100T)

### Data Management Capabilities (Filesystems management, authentication and authorization, distributed data systems) [~5 FTE]

- ❖ VPAC operates distributed sites at 4 locations, with major storage and compute resource across 3 of those sites (acting as a testbed for grid-enabled applications)
- ❖ VPAC manages authentication and access control and a national and internationally recognized Certificate Authority for the APAC Grid and its successor Pfc ICI
- ❖ Data set hosting

### Data Applications Development and Support [~6 FTE; J2EE and .NET]

- ❖ Web Portals and Content Management Systems (e.g, Plone, Drupal)
- ❖ XML (schema design, meta-data handling, validation, and transformation)
- ❖ Relational database design and deployment
- ❖ Image database applications, including geospatial data processing and integration (google-earth and ESRI tools)
- ❖ Tiered architecture design
- ❖ Meta data management tools

More information is available on our website ([www.vpac.org](http://www.vpac.org)).