

# **NCRIS Capability 5.16**

## **Progress Report**

### **Platforms for Collaboration**

Facilitator: Rhys Francis

Date: 14 November 2006

# Contents

1	Introduction	3
1.1	Context	4
1.2	Reference group and web site	5
1.3	Consultations	5
1.4	Service providers	5
2	National e-Research Infrastructure	6
2.1	Overview	6
2.2	NCRIS Investments	7
3	Progress with consultations	8
3.1	High Performance Computing	8
3.2	Advanced Networking	8
3.3	Expertise	9
3.4	Authentication and authorisation services	10
3.5	Data management services	11
4	Outline of Investment Structure	14
4.1	Overall Mission	14
4.2	Context	14
4.3	Potential Activities	15
4.4	Budget Issues	15
5	Project Plan	16
6	Expenditure to date	18
	Appendix I Reference Group	19
I.1	Membership	19
I.2	Terms of reference	19
	Appendix II Investment Principles	20
II.1	General Comments	20
II.2	Specifics	20
	Appendix III Research Information Infrastructure projects	21
	Appendix IV AARNet	43
	Appendix V APAC, the National Facility, the National Grid	48
	Appendix VI State based APAC Partners	52

# 1 Introduction

Modern research is increasingly dependent on technological platforms that enhance the research community's ability to generate, collect, share, analyse, store and retrieve information.

Some research can only be progressed because of the capabilities provided by these ICT platforms.

While termed "platforms for collaboration" in NCRIS, they are also described using the term "cyber-infrastructure"; and their application within research activities, and the consequential change in the way research may be carried out, is often referred to as "e-Research".

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

Within this context, five key areas were identified within the NCRIS roadmap<sup>1</sup>:

"Platforms for collaboration include the following sets of inter-related components:

- *Data storage management, access, discovery and curation* to improve interaction and collaboration;
- *Grid enabled technologies and infrastructure* to enable seamless access to the facilities and services required in various research fields;
- *Support skills* to assist researchers in developing and using this infrastructure effectively;
- *High performance computing* to allow analysis, modelling and simulation; and
- *High quality network access through high capacity bandwidth* to permit interaction with diverse data and computing resources."

The consultations for 5.16 and the other NCRIS capabilities confirm these as critical capabilities.

While undertaking the consultations, several statements of longer term intent (or goals) for an Australian national e-Research infrastructure have been suggested, statements such as:

- All networked research resources should be accessible through single sign on
- A common electronic access and authorisation framework should apply to all research data
- A common high speed network should connect all researchers and all research resources
- "Peta" scale (and beyond) capabilities should be available and made accessible

With visions of such scale, a number of issues present themselves within the context of planning within the NCRIS Platforms for Collaboration activity. Most notably:

- Institutions within the system will spend on the same activities, and spend far more than NCRIS, so that a clear understanding of the NCRIS role within e-Research infrastructure is important
- Underpinning components, such as researcher authentication and e-Research toolkits may also need to be considered for investment to ensure the infrastructure is easier to adopt and apply
- The inter-relatedness of components means that priority cannot be simply addressed by selecting some needs over others, a funding distribution will need to relate to some weight given to each need, and the corresponding readiness of solutions, technologies and associated products
- The state of readiness is also different across communities suggesting that support for adoption as well as delivery may be needed, and that the distribution of funding might change over time
- The funds available to NCRIS are expected to be insufficient to meet the overall need, so that extensibility of the investment structure is important to allow additional funds to be applied

---

<sup>1</sup> National Collaborative Research Infrastructure Strategy, Strategic Roadmap, DEST, February 2006.

## 1.1 Context

The context for the investment in Platforms for Collaboration is global in scale, and a variety of initiatives around the world provide guidance.

Broadly speaking this ground work has been recently and expertly covered by the e-Research Coordinating Committee (eRCC), some details of which are available through its discussion paper<sup>2</sup>, which says, inter-alia:

“Successful research is increasingly team-based. It is also increasingly necessary for research to be carried out across disciplines and across geographic boundaries, as researchers attempt to address more complex issues where boundaries are less relevant.

Developments in information and communications technologies (ICT) are enabling large amounts of data to be manipulated and transferred very quickly across long distances on advanced networks. Developments in ICT are also changing research methodologies and enabling formerly inaccessible problems to be addressed. And in both ways, ICT developments are enhancing our ability to approach complex problems.

A number of countries are taking significant steps to enable researchers to utilise advanced technologies. Many researchers in Australia, whether in universities, research organisations or industry, have access to advanced computing and network capabilities linking places in Australia and overseas. However, as ICT develops very quickly, Australian researchers will require specific support to exploit the full potential to share data and information and work more collaboratively. “

and

“Status of Australia's e-Research Infrastructure

A focus on excellence is critical to ensuring internationally competitive outcomes from research and the high returns that can arise. The nature of excellent research is increasingly characterised by collaboration on a national, and more often, a global scale.

To conduct top quality research, Australia needs to maintain a world competitive electronic research infrastructure (e-Research Infrastructure) and have researchers who can use that research infrastructure in an increasingly sophisticated way.

...

Australia's strategic investment in e-Research infrastructure is consistent with similar initiatives in leading research communities, such as cyber-infrastructure in the US, e-Science in the UK, European e-infrastructure and GRID Canada.“

The eRCC committee recommended a range of activities should be supported, only some of which might meet NCRIS requirements as ‘infrastructure’ and which were intended in any event to be in addition to infrastructure investments. Therefore while 5.16 will ensure its investments align with the directions and intent set out by the eRCC, the activities the eRCC proposed will need separate funding.

As well as the views developed and represented by the eRCC, the development of the 5.16 investment plan intends to align with international developments and be explicitly informed by:

- the needs of NCRIS investments which are summarised in a companion report from the facilitator<sup>3</sup>
- the discussions and recommendations of the PMSEIC working group on Data for Science
- consultations in data management, authentication and authorisation, and e-Research toolkits
- reviews and inputs from the existing networking, high performance computing and grid service providers

---

<sup>2</sup> An e-Research Strategic Framework, A Discussion Paper, DEST June 2005

<sup>3</sup> Investment Plans and Platforms for Collaboration, DEST, October 2006 (see [pfc.org.au](http://pfc.org.au))

## 1.2 Reference group and web site

A reference group has been established and has met three times, and is expected to hold a further three full day meetings over the remaining course of the planning activity. The terms of reference and membership is given in Appendix I. A number of email lists and a web site have been established to support the development of the investment plan (see [www.pfc.org.au](http://www.pfc.org.au) ).

## 1.3 Consultations

The investment plan is being informed by a number of specific consultations and several activities which were initiated separately to the 5.16 consultations, as detailed below.

The report of the e-Research Coordinating Committee	Provides an assessment of the expertise and assistance research groups need to effectively adopt the developing e-Research infrastructure
PMSEIC Data for Science working group	Provides an agreed strategic view of relevant issues and the associated systemic requirements
NCRIS investment plans	Provides requirements mostly related to data management and grid based collaboration systems
DEST review of APAC	Provides an assessment of the need for supercomputing, grid deployment and data transport services, and the role APAC might play in delivery of those

Documents related to these are available separately, although the PMSEIC Data for Science working group report will not be available until after PMSEIC meets in December.

Discussions related to the NCRIS investment process have led to a set of principles for 5.16 investments which are provided in Appendix II.

Early work also identified a range of additional issues which would need to be explored. These relate to the support needed within a national framework for research collaboration that permits:

- Institutions to easily recognise each other's research staff
- Facilities and data to be easily accessed, while supporting appropriate security, privacy and confidentiality requirements
- Researchers to easily locate and access resources relevant to their research goals

Broad based consultations are therefore in progress as follows (see [pfc.org.au](http://pfc.org.au) for more details):

- AAA – to identify a trust federation that could be permanently established and operated to support identification for research as well as other higher education institutional purposes
- Data management – to identify the categories of data and data management missions, so that systemic requirements can be identified and a role for NCRIS investment agreed
- e-Research toolkits – to identify the state of play in the tools and technologies that researchers might deploy to take advantage of e-Research infrastructure (from the desk top)

## 1.4 Service providers

In addition to formal interactions with AARNet and APAC, additional investigations are planned or already in progress with other participants around their ability and interest in supporting relevant services and service delivery requirements, including:

- National institutions: the Bureau of Meteorology, CSIRO, ANU, Geosciences Australia and the National Library of Australia
- Regional institutions: the state based APAC partners, research intensive institutions, and e-Research activities; to ascertain the extent of existing services and development plans
- Other service providers: such as Aus-cert and a variety of "collections" agencies

## 2 National e-Research Infrastructure

### 2.1 Overview

At present, e-Research infrastructure service delivery is dominated at the national level by AARNet (expending circa \$40M pa) and APAC (expending circa \$20M pa).

However, this picture misses the interests of other key participants. For instance:

- The Bureau of Meteorology (BoM) and CSIRO jointly invest more than \$10M pa in HPC and data services within the High Performance Computing and Communication Centre, independently of APAC, and both invest substantially in related infrastructure outside the HPCCC
- Many of the largest data holders in the country, Geoscience Australia, BoM, many state and federal research agencies (such as in the primary industries, marine and health areas), and the humanities overall, are not perceived to be connected to a common national e-Research infrastructure
- Substantial and ever increasing data management capabilities exist around the country, including at least five petabyte scale near-line storage systems, and nearly all future research data will be born digital, and yet a co-ordinated approach to national data management has not arisen
- Several regional research networks now exist, creating a federated network model
- DEST initiated Research Information Infrastructure projects representing an annual spend of \$17M (average over last three years) are only weakly connected with the service providers and there is insufficient support for transitioning developments into products and services
- A shibboleth trust federation has been created but with uncertain future while significant nation wide infrastructures, such as the APAC National Grid and CSIRO's enterprise management services, remain separate to this trust federation
- Many state governments intend to invest in e-Research activities and yet the co-ordinated development and delivery of services against the needs of research groups lacks a home
- The APAC state partners spend an additional \$20M pa on activities historically identified as outside APAC's core interest, activities which then remain uncoordinated at the national level
- NCRIS capabilities themselves will spend an estimated \$10M pa on e-Research infrastructure within the various capabilities, with no available means of co-ordination
- LIEF grants are in excess of \$30M pa with some component in e-Research infrastructure, and Universities themselves invest in eResearch infrastructure, all of which is carried forward independently and without strong coordination

These points appear to show a system developing in piecemeal fashion, with weak co-ordination and few anchor points for research communities; there is certainly no 'one-stop shop' for developing and supplying e-Research solutions to meet the needs of a research community or a research facility.

To provide a view across the national space, some capability and summary statements are appended:

- Appendix III provides a summary of the current Research Information Infrastructure projects
- Appendix IV contains a summary capability statement for AARNet
- Appendix V contains capability statements from the state based APAC partners
- Appendix VI contains a summary of APAC, the National Facility and the National Grid

Overall, awareness of the need for co-ordination and the importance of standards is growing. The policy barriers to collaboration and the impact of uncoordinated investment are less well understood.

Finally, the aggregate e-Research infrastructure expenditure in just the entities named above exceeds \$100M pa, and of course all research institutions manage major (and in aggregate much larger) ICT budgets, the components of which need to combine with and become part of the overall e-Research infrastructure.

Thus, as the vision for e-Research infrastructure hinges on inter-operability, universal access, coverage, and strategic coherence; investment priorities for Platforms for Collaboration will need to support improved leadership and improved co-operation as well as providing key shared services.

## 2.2 NCRIS Investments

A detailed report is available<sup>4</sup> covering the implications for platforms for collaboration that arise from the proposed NCRIS investments. Highlights from the report are as follows.

### e-Research infrastructure investment

About \$10M pa of NCRIS funds in the investment plans may be being directed towards specific elements of e-Research infrastructure. The activities within that \$10M pa are mostly focussed on data access and platform integration across multiple sites, with little in modelling, HPC or networking.

Also, aggregating with 5.16 yields a funding rate about half that of the SII funding rate over the last three years. The result may be a reduced rate of general e-Research infrastructure development and significant difficulty in providing additional services

### Data Management

Overall significant attention has been paid to the capture, curation, and access, for “community” data.

In general the investments leave responsibility for data with the facility (5.1, 5.2-phenomics, 5.10) or with the institutions related to collections (5.2-atlas, 5.12, 5.13) and the researcher otherwise (so that data is then managed according to the work practice of the researcher’s environment).

### Grid

The requirements for different grid-like capabilities can be distinguished as follows:

Grid service	5.1	5.2	5.3	5.4	5.5	5.8	5.10	5.12	5.13
Integrated multi-site platform	X	X	X						
Distributed shared view of data	X	X	X			X	X	X	X
Remote access to compute power	X	X	X	X	X	X	X		X
Real-time multi-site data analysis							X		
Virtual presence and control			X				X		
Access Grid like collaboration				X	X	X		X	X

### Expertise

A significant part of the this expenditure will relate to e-Research expertise, which will be in short supply. Mechanisms to grow and share key human resources need to be provided.

### High Performance Computing

The overall needs of modelling and theoretical research, and the increased demand that may arise from research activity associated with the NCRIS facilities is likely to easily exceed supply.

The benefits of speed vs capacity and the best means of user engagement need to be investigated.

### Networking

A desire has emerged for subscription based research traffic from all NCRIS funded facilities to all Australian researchers; from shared facilities through to every researcher’s desktop.

As investments are made, the network backbone is likely to become less homogeneous (some links with higher speeds), so that a revised pricing framework may need to be developed and agreed.

<sup>4</sup> Investment Plans and Platforms for Collaboration, DEST, October 2006 (see pfc.org.au)

### 3 Progress with consultations

Consultations around e-Research Toolkits and the broader concept of the grid have yet to take place.

#### 3.1 High Performance Computing

Any given researcher and research community will have access to a variety of compute and storage capabilities, including large systems operated on a merit basis; mid-range systems at institutions operated on a shared access basis; and smaller and often dedicated departmental clusters and desktop systems. Consequently, the HPC needs of a research community are difficult to estimate, and the degree to which central or co-ordinated provision is relevant, is even more difficult to estimate.

However, the comments on HPC made during the APAC review provides several observations, perhaps most importantly related to the scale of investment.

- Additional expenditure is required if the peak facility is to be retained at the historical level
- A more frequent purchase rate is needed to improve the return in Tflops delivered against dollars, under the assumption that a significant overlap in the operational periods of systems is manageable within the machine room infrastructure
- Given the already competitive nature of access, more resource overall should be provided to allow for the broader clientele envisaged under NCRIS

The different requirement for rate of computing (peak Tflops/s or capability) versus the amount of research computing that can be supported (Tflop-years or capacity) is left unclear.

While the annual average for both, and the corresponding rates per dollar spent, are all significantly improved if the frequency of purchases can be increased, which of the two measures one is aiming at dominates the strategy for provision.

Overall, some of the implied need associated with the NCRIS investments may be able to be met by significant capacity rather than requiring peak capability, however this is untested.

#### 3.2 Advanced Networking

The NCRIS investments suggest a generalisation of the existing research network that needs to be better understood, both in terms of detailed requirements and in provisioning options.

We can note overall that NCRIS investments:

- include many examples of grid like systems providing shared access to infrastructure operating across multiple sites with a national distribution
- includes in 5.3 and 5.10, examples of very high real time data generation rates that may require special analysis and treatment
- represent a substantial move to develop federated views of widely dispersed data (in all but 5.4 and 5.5), which may lead to service level requirements designed to support rapid search and access

The manner in which the research network can be generalised to encompass a range of government agencies and state based research institutions needs to be determined as does the manner in which costs can be met by volume and destination independent subscription charges.

Also, the actual user experience in access to the research network, depends ultimately on institutional infrastructure and institutional policies, that may also need to be considered. Apart from 5.3 and 5.10, each platform proposed under the NCRIS investments might reasonably operate over a quality of network connectivity that could be met with 1Gbps tails. How that reaches sites on campus and research desktops in general needs to be further discussed. Also, to support multiple investments with such connectivity, some of the research intensive institutions will need significantly higher aggregate connectivity which may also create network issues within those institutions.

### 3.3 Expertise

One of the three commonly identified issues in moving towards a more coherent development of e-Research infrastructure concerns the development and provision of appropriate expertise.

An analysis of investments in e-Research infrastructure across the NCRIS capabilities, identifies a level of spending that will ultimately translate in to a demand for expertise. The investment plans themselves show that the various communities are at different stages of development towards an e-Research perspective; which means they will necessarily have access to very different levels of such expertise. This will be true more broadly.

Also, e-Research involves the use of multiple and entirely unrelated specialisations, such as curation of data, advanced networking, or parallel software for supercomputing. Added to this, grid capabilities and middleware are a rapidly evolving set of specialisations in their own right (such as searching, authentication and authorisation). Research groups cannot possibly cover this space.

Some important factors related to e-Research expertise are:

- Expertise management is enhanced by building groups of specialists rather than relying on unrelated individuals
- Different expertise and different levels of expertise are required during different stages of a communities migration toward e-Research
- Eventually some expertise needs to be embedded in communities (eg. data curation) and some needs to be embedded in service providers (eg. network management)
- Along the way, flexible collaborative teams are needed so that the infrastructure can evolve as the requirements are better understood

Also, as communities become more e-Research oriented, they tend to co-evolve services for data generating and gathering, with services for information analysis and re-use. This happens because each community needs to develop a consensus on the standards required for inter-operation.

Standards development is always a long iterative processes, which means that researchers will undertake bespoke software development in order to continue their research within the evolving context. Further difficulties then arise as reliance is placed on such software leading to a need for improved software engineering and particularly software productisation expertise.

Many of these issues were identified by the e-Research Co-ordinating Committee and the basic perspective developed by that committee remains valid and is reflected here.

A solution to expertise is beyond the budget of Platforms for Collaboration and a broadly based consultancy activity is most likely out of scope from an infrastructure investment point of view.

However the problems arising from the fragmentation of e-Research infrastructure through divergent directions in investments is exacerbated by this missing expertise and in particular the leadership that could be expected from high levels of expertise.

Consequently, some actions need to be taken and the investment plan for 5.16 cannot be silent on expertise and leadership.

- Some form of recognised peak body or association needs to be formed, the exact role needs to be resolved, but some initial suggestions are made in a later section.
- Investments in 5.16 need to be structured to enhance and provide access to pools of expertise where they can be found or where other parties will co-invest

It should also be noted, as a rough estimate, perhaps 50 new experts in various areas of e-Research infrastructure may need to be appointed into NCRIS platforms over the next five years, in addition to expertise developed within service providers.

### 3.4 Authentication and authorisation services

Another of the three commonly identified issues in moving towards a more coherent development of e-Research infrastructure concerns authentication and authorisation. A uniform approach to authentication is critical if ad hoc collaborative groups are to be easily supported.

- Any consideration of national e-Research infrastructure leads immediately to the need for a shared method of identifying researchers across institutions that eliminates creating new identities, new logins and new passwords every time a person is added to a project or a collaborative activity
- A second need then appears around systems that allow owners of resources to say who can do what; and to define such authorisations for access in terms of roles and categories of people
- The third component can then be identified that relates to the provision of access control systems that enforce those decisions reliably and with certainty within applications
- However, before any of this can be deployed systemically, a range of framework and policy decisions need to be made, for which community agreement is needed, so that trust is established and the system that provides authorisation is easy to use

The consultation process within Platforms for Collaboration brought together representatives from the Open Access to Knowledge Law Project, the Middleware Action Plan and Strategy project, the E-Security Framework for Research project, the Meta Access Management System project, the APAC National Grid, AARNet and CSIRO, to review what was known and what plan might be developed.

Agreement has been reached on a path forward noting that the technology and product support for each of the needs identified above is in different stages of development and yet all need to be in place for acceptance of the system to be wide spread. Overall:

- A trust federation will be established to provide the foundation on which identity and authorisation servers and transactions can rely. This trust federation will at least cover the entire Higher Education and Research sector, and may federate with other national trust systems where possible, such as may be implemented for various government agencies
- The trust federation will support PKI and shibboleth, and may support other standards as required
- To support the formation and ongoing operation of the trust federation, discussions will be convened within CAUDIT to consider the approach to compliance that should be taken, in terms of policy, framework, escalation, advice, remediation, and training
- Effort within the projects listed above will be directed towards developing a range of next steps, including: a Shibboleth federation policy; a suitable minimum set of attributes and mechanisms for extensions; bridging between the APAC National Grid and the trust federation; the development of some simple use cases; and the working up of the required directories and policy enforcement components to support those use cases

How the trust federation will be structured legally and organisationally needs to be determined.

Such a trust federation is a key system service essential for collaboration, and as it will simplify and support operational aspects of the various platforms and all of the access regimes envisaged by the NCIRS capabilities, it appears a natural fit for investment by 5.16.

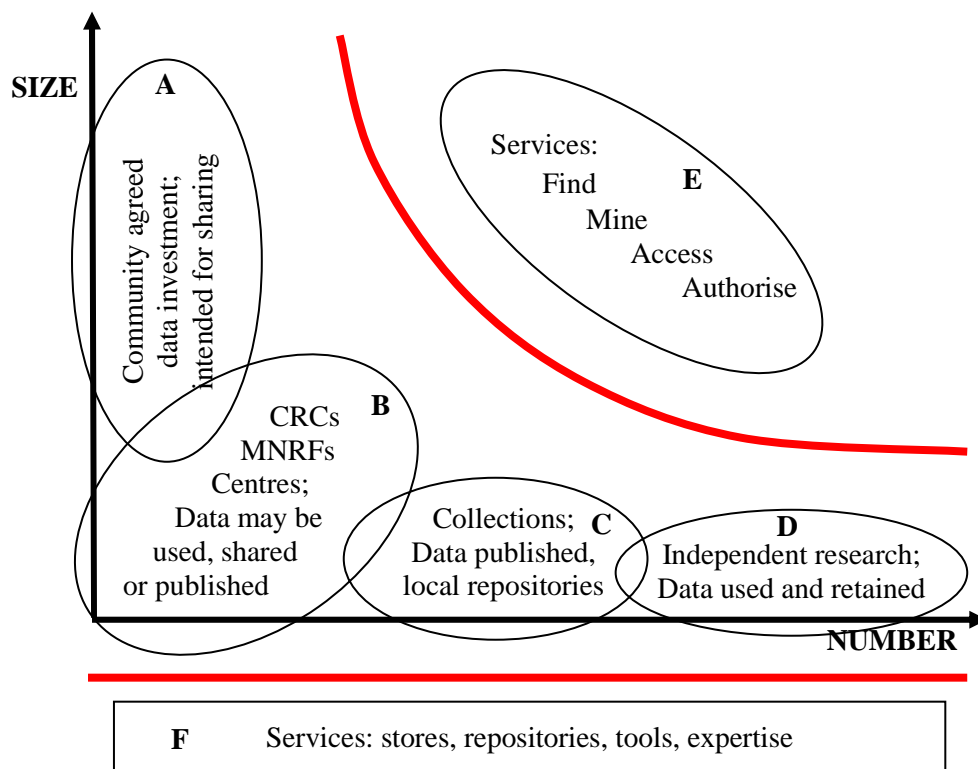
However the trust federation will also be able to support services related to a range of non research activities (such as student and staff resource access) and may also federate to non research organisations to assist resource sharing, suggesting either some form of subscription funding may be appropriate or that funded value adding services should recoup some of the base level operational cost.

The participants in the AAA consultation intend to provide a detailed plan with governance, budgets, funding model and service roll-out within the 5.16 Investment Plan.

### 3.5 Data management services

The last of the three commonly identified issues in moving towards a more coherent development of e-Research infrastructure concerns data management.

Consultations and surveys undertaken by Platforms for Collaboration suggest that data volumes are growing rapidly and that a significant fraction of research data is held in isolated forms and not easily accessible, and that a number of different missions exist around data, depicted below.



Within these cases, two kinds of data can be distinguished: private and public data. Private data is usually held by a researcher through self generation, or by acquisition from surveys, an instrument or device. Public data is data which has been placed in a public repository for general public access. Often, researchers also make some of their private data available on a controlled access basis, e.g. by copying to collaborators, but do not normally provide public access to all the data they hold.

In some disciplines where data is derived from significant public or shared investment, a common data holding protocol applies in which each researcher is granted exclusive or controlled access to the data collected on their behalf for a limited time. After that time the ownership of the data reverts to the institution in which the researcher is based, or is made public through a repository.

With reference to the diagram, for the different missions, different properties and responsibilities are present, and the role of NCRIS needs to be determined.

- A. Several communities have established practices that capture and share data of general value. These tend to form large datasets that are generated or gathered as the result of significant investments; such as in astronomy, high energy physics, earth observations and bioinformatics. The size of these data sets tend to be very large, typically in terms of tens to hundreds of terabytes and sometimes petabytes, with high growth rates.
- B. At a slightly smaller scale and more numerous, various organisations e.g. CRCs and MNRFs, have the gathering or generation, curation and publication of data as part of their mission; or use data for internal purposes, the results of which may be published in the form of information services; or they provide research support services that generate data. The data often is privately held, and even when it is published, it is in the form of processed data, the raw data may not be made

available to the public. The size of these data sets tends to be in the many gigabytes to terabyte range.

- C. Continuing to decline in scale and increase in number, many research organisations, departments, teams, and even individuals have established data collections the contents of which are intended for publication and access. These publicly available data sets (donated by researchers) are normally kept in institutional or personal archives or web sites. The size of these data sets tends to be in the multiple gigabyte range.
- D. At the end point of this curve, nearly all researchers generate data or store data on the desktop, much of which is only ever intended for individual use, and much of which is inaccessible to other researchers and only visible through the publication of derivative research results. Most of this raw data, would have to be uploaded to a web site or institutional repository to make it accessible. The size of such data sets can be quite variable depending on the devices or computational models that generate the output.
- E. A separate class of mission relates to organisations that seek to provide access to research data, either within disciplines or more broadly, and which may or may not hold the primary data themselves. These organisations provide catalogues, abstracts or thumbnails, and can search and possibly mine data over which they hold indexes. These sources can only provide access to public data or private data researchers have decided to make publicly accessible. The aggregate size of data accessible in this way would be in the range of terabytes to petabytes.
- F. A final class of mission relates to service providers seeking to support the retention, curation, access and analysis of data sets. Generic service providers necessarily operate only on publicly accessible data, although some specialised services could operate on controlled access data. Their added value lies in the co-location of data sets and the consequent ease of datamining and linking that becomes possible. Many institutions might operate a similar service on their own behalf for their retained private data. The size of repositories tends to be in the range of hundreds of terabytes to many petabytes, and exabytes would be conceivable in some settings.

This discussion highlights that data management services will be provided by a variety of sources and the development of investment decisions for data management by NCRIS in general and 5.16 in particular needs to be based on principles that fit within a broader framework.

Fortunately such frameworks are under development, and 5.16 planning can make the following assumptions in line with current thinking by the AVCC, NH&MRC and ARC.

Institutions should:

- Develop and implement a policy on data ownership
- Provide guidelines to researchers on ownership, what to keep and researcher responsibilities
- Maintain durable records on what research data has been held and ensure that research data is under the control of the institution where the work was performed
- Provide secure systems for holding data and for granting access to that data

Researchers should:

- Determine what data to keep, considering research community practice and any project or legal requirements
- Ensure research data is retained (for at least 5 years from publication of results) using institutionally provided mechanisms
- Ensure at the end of employment (for whatever reason) data retention passes to the institution
- Maintain confidentiality where it exists

Also, data actually moves through a life cycle, and gains value from the process, so that the use of the term data here is broadly based to include digital objects.

In consideration of the above, the following general principles will be applied in determining the services and support 5.16 might provide.

- Investment by 5.16 should be directed towards data which is of value and preferably lasting value and available for public access, either immediately or ultimately
- In general, investments (by other sources) intended to support data acquisition should be expected to also fund or arrange retention and access services for that data for the lifetime of the investment
- Retention of privately held data or data to which access is highly limited is the province of the institution(s) sponsoring the research

Taking these principle into account we arrive at some possible investment goals.

- 5.16 could support public data retention when investment periods expire, where that data is important or irreplaceable; across all categories in the diagram above. However the permanent preservation of data from category A is a difficult problem given the potential scale, but is also not expected to arise within the current investment horizon for 5.16
- 5.16 could provide the means by which collections of public data (generated in categories C and D) could be preserved and made accessible, on the basis that such a service adds considerable value to the data and will not otherwise be easily or quickly provided. This would relate to data specifically nominated for public access and deposited by researchers into public repositories
- Such a “collections service” could be extended to also include public data deposited into repositories from activities in category B
- Of the two sets of services depicted in the diagram, those related to re-use (E) and those related to retention (F), support on the re-use side (E) is clearly systemic, beyond the bounds of individual institutions, complementary and value adding to individual data holdings
- Therefore 5.16 could focus on data search, data mining and access control services in E and perhaps national data location and movement services in F

There is another issue which may need to be considered in developing the investment plan, that being the benefits of co-location of data and compute resources. At present network bandwidth is perceived to be sufficiently high to transport data to compute resources. However, the ever increasing amount of data may make it more efficient to install processing power close to the sources or repositories of data. This suggests a future intention to build towards a number of e-Research support facilities that provide for the co-location of compute and data storage resources, is a desirable goal.

Underlying this analysis is a governance question, namely which body will have the authority to decide on the longevity of data sets, particularly if they are considered to be “nationally significant” and worth long term archival. Clearly any funder can do so for the duration of their funding, the issue arises when investments expire and some transfer of responsibility is needed. In addition, sponsors of research will need to be clear which data generated from their sponsorship is private, limited access, immediately public or ultimately public.

These issues, along with the implications of the stages in the data life cycle as well as requirements for supporting infrastructure such as universal identifiers will be discussed in further workshops.

## 4 Outline of Investment Structure

The Australian national e-Research infrastructure can be expected to take the form of a combination of national and regional activities due to the decision making processes of governments and institutions.

A foundation principle for the investment plan is to strengthen and build on co-operative arrangements so that an increasingly coherent level of support can be provided to researchers, and their collaborations and communities.

The goals of the investment plan for 5.16 are to provide e-Research infrastructure services which are:

- of broad value across all research communities
- of value to those communities whose research needs led to the NCRIS capabilities
- of value to the services delivered by the platforms created or augmented by NCRIS investment

Hence 5.16 investments will focus towards shared services that no single community could afford or justify in its own right, or where piecemeal development would hinder rather than aid collaboration.

The intention is that such services should be delivered through agencies where the service is the primary mission, so that over time, other capabilities and research investments can source (potentially dedicated) services from this infrastructure and provide appropriate funding to it for that purpose.

At a functional level, this demands a highly co-operative approach across suppliers of infrastructure and services so that the user experience is more cohesive in terms of data management and access, the sharing of compute capabilities, the controlling and setting of authorisations, and an ease of remote interaction through readily available “always-on” collaborative environments.

### 4.1 Overall Mission

The term eRI, for e-Research Infrastructure, has been adopted herein as a place holder for this overall goal and the entity which might occupy that position. That entity could be a development of the existing APAC partnership or it might be a new entity, the governance arrangements are deliberately left as a subsequent discussion at this time.

eRI is proposed as the peak body tasked with developing and sustaining the forums in which co-operation can be achieved and which over time provides policy and standardisation frameworks that deliver a nationally coherent e-Research infrastructure.

Its mission will be to identify, develop and deliver nationwide and world class services and expertise, that support effective e-Research within and across all research disciplines.

This will include services and expertise related to:

- data capture, management and retention
- data publication, discovery and re-use
- data analysis
- computational modelling
- collaboration systems
- grid inter-connectivity
- networking

### 4.2 Context

The development of e-Research and related infrastructure requires a mixture of:

- Policy advice and implementation
- Consulting to bring together requirements and solutions that meet user needs within sustainable architectures and deployable services

- Provisioning of services and directing service development through service providers

The stakeholders relevant to eRI include researchers and communities of research, research facilities, research support services, institutions and collaborations of institutions, and their funders.

The manner in which these entities will be included within the governance of eRI and/or the service providers funded by or related to eRI will be considered at a later stage in the planning.

### 4.3 Potential Activities

The following activities represent a core set, further work may identify additional services or responsibilities within proposed service areas.

#### e-Research Infrastructure (eRI)

- Providing a framework for investment and allowing for further activities and funding
- Promoting e-Research and the deployment and use of e-Research Infrastructure
- Mapping user/application requirements, priorities and expectations, to assist services meet needs and transform the landscape
- Providing policy advice regarding e-Research deployment and adoption

#### e-Research Consulting (eRC)

- Developing and deploying staff within eRI and from other relevant entities (service providers and user communities) in a co-ordinated and expert e-Research consulting service
- Providing a team solution to researcher, research community and research facility support

#### The e-Research Trust Federation (eRTF)

- Providing the central support (including auditing) for PKI and shibboleth infrastructures

#### The Australian National Data (Collections) Service (ANDS)

- Providing retention services for important collections
- Hosting a range of access, location, and datamining services

#### The Australian National Grid (ANG)

- Supporting access and interoperation for compute and data services
- Hosting instrument/facility/data integration to the grid
- Providing advanced collaboration and visualisation environments

#### The Australian National (Computing) Facility (ANF)

- Providing a peak HPC/data capability service, accessed by shares and merit allocation

#### The Australian Earth Systems Science Facility (AESSF)

- Providing share based access to a peak capability service for Earth System Science research

#### The Australian Research and Education Network (AREN)

- Providing a broader, more capable, subscription based research and education network

### 4.4 Budget Issues

The total current activity in these areas (including activities funded through SII relevant to AAA) represents about \$35M pa from DEST and a further \$65M pa from participants. In addition, the first two items have a new breadth and are therefore not currently supported within that funding.

The NCRIS provision of \$15M pa, if the sole funding line, requires participants to take a greater share of the burden, or implies a reduction in activity. The intention in further planning is to find ways to support all of the above activities if possible.

## 5 Project Plan

Month	Activity	Details	Changes	Responsible
May, June	Start-up	Agree reference group, role and scope, contract, engage DEST/facilitators		Rhys, Ian
July	<a href="#">R/G Meeting #1</a>	Establish role and scope, identify actions, review 1 <sup>st</sup> deliverable		Rhys, DEST
23 Jul	<a href="#">1<sup>st</sup> deliverable</a>	PfC implications from Progress Reports		Rhys
August	5.1 IT Architecture Activities <a href="#">R/G Meeting #2</a> Networking BoM/ACCESS	Engage help for review of IT requirements in 'omics Scoping work around proposed workshops Implications of initial NCRIS feedback, scope tasks Establish activity to provide network input to the plan Establish activity to include BoM/ACCESS input to the plan	Delayed	Rhys, Ian, Paul Davis Reference Group Reference Group Rhys, Ian Rhys, Ian
September	APAC review Networking 5.1 IT Architecture 5.8 Collab scoping AAA workshop #1 Collab workshop#1 Archer scoping	Engage reviews and results Progress information from aarnet Finalise outline of 'omic IT requirements, trial with IT Directors Assist develop scoping for bio-security collaboration environment Bring parties together to gain views on 1, 2, 5 and 10 year visions Bring parties together to gain views on 1, 2, 5 and 10 year visions Outline information management needs across capabilities	Delayed     Replaced	Rhys Rhys, Peter, aarnet Paul Davis Markus Buchhorn Daniel, Nick, et al Daniel, et al Ah Chung, et al
October	<a href="#">R/G Meeting #3</a>	Review Progress Report and associated inputs (as above)		Reference Group
6 Oct	<a href="#">2<sup>nd</sup> deliverable</a> Data workshop #1	Provide written comment on the first 9 Investment Plans Bring parties together to gain views on 1, 2, 5 and 10 year visions		Rhys Linda, et al

NCRIS		Platforms for Collaboration	Progress Report	
19-20 Oct	Network Review #1	Bring parties together to gain views on networking requirements and options	Delayed	Peter, et al
30 Oct	<a href="#">3<sup>rd</sup> deliverable</a>	Attend NCRIS Committee meeting as required		Rhys
		Progress Report including overview of existing infrastructure	Delayed	Rhys
November	AAA workshop #2	Develop content for path forward	New	Daniel, Nick, et al
	Collab workshop #2	Develop content for path forward	Replaced	Daniel, et al
	Data workshop #2	Develop options for investment related to data management	New	Linda, et al
	Archer workshop #2	Consolidate information management options across capabilities	New	Ah Chung, et al
	NCRIS Investments	Consolidate requirements from NCRIS investments		Rhys
	Service providers	Discuss options for service supply, aarnet, apac, others		Rhys
December	<a href="#">R/G Meeting #4</a>	Review outputs from work to date, determine any additional activities		Reference Group
	eRI Members	Scoping governance arrangements with eRI members	New	Rhys
	e-Research Toolkit #1	Consider what tools and technologies could be promoted to assist e-Research uptake	New	Rhys, Daniel, Paul et al
January				
February	AAA workshop #3	Finalise proposal for trust federation		Daniel, Nick, et al
	Data workshop #3	Finalise proposal for any data management investment		Linda, et al
	e-Research Toolkit #2	Finalise any initial tools and technologies to be supported	New	Rhys, Daniel, Paul et al
	<a href="#">R/G Meeting #5</a>	Review draft investment plan		Reference Group
23 Feb	<a href="#">4<sup>th</sup> deliverable</a>	Draft Investment Plan		Rhys
March	<a href="#">R/G Meeting #6</a>	Provide DEST with input on Investment Plan		Reference Group
		Make presentation to NCRIS Committee		Rhys
30 Mar	<a href="#">5<sup>th</sup> deliverable</a>	Revised Investment Plan and Final Report		Rhys

## 6 Expenditure to date

A summary of expenditure to date against the original budget headings along with an estimate to the end of the consultation process is provided below.

9-Nov-06

NCRIS - Platforms for Collaboration

LIFE TO DATE FINANCIAL SUMMARY

01 May 06 - 31 October 2006

	<b>Budget</b>	<b>Total to date</b>	<b>Estimate</b>
<b>Income</b>			
DEST funding	250,000.00	100,000.00	250,000.00
<b>Total Income</b>	<b>250,000.00</b>	<b>100,000.00</b>	<b>250,000.00</b>
<b>Operating Expenditure</b>			
<u>Facilitator Expenses</u>			
Remuneration/on-costs	\$157,531	\$50,340	\$157,531
Administration support	\$6,922	\$2,141	\$6,922
Travel	\$23,000	\$8,543	\$22,543
Incidentals	\$2,000	\$245	\$1,145
<u>Reference Group meetings</u>			
Travel	\$24,000	\$3,724	\$11,245
Accommodation	\$9,000		
Incidentals	\$3,947	\$458	\$1,458
<u>Activities &amp; Consultations</u>			
AAA Workshops	\$7,200	\$1,488	\$4,488
Data Workshops	\$3,600	\$3,447	\$6,447
Collaboration Workshop	\$7,200	\$0	\$4,000
Network Review	\$3,600		\$0
Archer Workshop	\$0	\$831	\$831
IT Architecture (5.1-5.13)	\$0	\$0	\$0
<u>Miscellaneous</u>			
Binding/Printing	\$2,000	\$0	\$2,000
<b>Total Expenditure</b>	<b>\$250,000</b>	<b>\$71,217</b>	<b>\$218,610</b>

## Appendix I Reference Group

### I.1 Membership

- Tom Cochrane, DVC(Technology, Information and Learning Support), Division of Technology, Information and Learning Support, QUT
- James Dalziel, Professor and Director, Macquarie E-Learning Centre Of Excellence (MELCOE), Macquarie University
- Peter Nissen, National Broadband Adviser for Education
- Linda O'Brien, Vice-Principal (Information), the University of Melbourne
- John Sheridan, Chairman, Deputy Dean, Department of Mechanical Engineering Faculty of Engineering, Monash University
- John Shipp, University Librarian, University of Sydney
- Nick Tate, Director, ITS and AusCERT, the University of Queensland
- Ah Chung Tsoi, Director, e-Research Centre, Monash University
- Nigel Ward, Executive Director, Australian ADL Partnership Lab

### I.2 Terms of reference

- The role of the reference group is to support the Facilitator to develop the investment plan for the NCRIS Capability, 5.16. The expected timing for this development is a progress report by November and a final report by March 07.
- The scope of the investment plan, as identified in the NCRIS Roadmap, includes: data storage and management, data access, discovery and curation; grid enabled technologies and infrastructure; support skills; high performance computing; and high capacity bandwidth.
- The purpose of 5.16 is to support the whole research community, not just those areas relating to Roadmap capabilities. The investment plan will seek to build on work undertaken within the Systemic Infrastructure Initiative.
- The investment plan will be specific about how NCRIS money should be spent to the end of 2007/2008. It may be less specific in out-years where it is considered desirable to leave flexibility for decisions to be made closer to the time of implementation.
- Reference Group members will support the development of the investment plan in areas relating to their particular expertise. The Reference Group will also advise on the needs of the social sciences and humanities, whose needs were identified in the Roadmap as falling most appropriately into this Capability.
- Development of the investment plan will be informed by consultations with key suppliers of services including AARNet and APAC. The Reference Group's considerations will be expected to take into account the outcomes of the review of APAC to be conducted in the second half of 2006 and the investment plans of the other capabilities
- After the initial investment plan is lodged, the reference group will provide a foundation for ongoing advice to NCRIS on issues related to this capability.

## Appendix II Investment Principles

This statement of principles has arisen from discussions around inter-relationships between 5.16 investment and other NCRIS capabilities.

### II.1 General Comments

The Goals of the investment plan for 5.16 are to provide e-Research infrastructure services which are:

- of broad value across all research communities
- of value to those communities whose research needs led to the NCRIS capabilities
- of value to the services delivered by the platforms created or augmented by NCRIS investment

Hence 5.16 investments will focus towards shared services that no single community could afford or justify in its own right, or where piecemeal development would hinder rather than aid collaboration.

The NCRIS Roadmap identified a number of likely 5.16 investments, including: networking, HPC, national grid services, the flexible provision of various forms of expertise and data related services.

The intention is that such services should be delivered through agencies where the service is the primary mission, so that over time, other capabilities and research investments can source (potentially dedicated) services from this infrastructure and provide appropriate funding to it for that purpose.

The geography of service providers and facilities will arise from co-investment and other discussions.

### II.2 Specifics

Storage	<p>5.16 intends to invest in storage, data movement and data linking services, within a national integrated system, supported by a number of suitable service focused organisations.</p> <p>It also intends to invest in expertise to help facilities join the national data backbone, and plans to provide data location and transport services to access and move data efficiently and effectively.</p> <p>5.16 does not expect to purchase storage systems located in and dedicated to research facilities or research groups.</p>
Computing	<p>5.16 intends to invest in a range of computing services that can be redeployed across a broad range of research users, supported by a number of suitable service focused organisations.</p> <p>It also intends to invest in expertise to assist facilities and research communities apply grid solutions to access these shared resources.</p> <p>5.16 does not expect to buy computers located within and dedicated to research facilities or research groups.</p>
Expertise	<p>5.16 envisages that some expertise related to the development of e-Research infrastructure and its adoption by facilities and research communities will be re-deployable and funded by 5.16</p> <p>5.16 envisages other expertise to be a 'cost of operation' of a facility (and hence out of scope) because the expertise is either so domain specific that it rests entirely within the needs of a specific facility or the personnel are to be full time and permanently required by a facility.</p>
Connecting	<p>5.16 intends to budget with co-investors for the full cost of the e-Research infrastructure services it supports and may invest in some access systems. However, costs associated with modifying enterprise systems within facilities to inter-operate with the e-Research infrastructure is out of scope.</p>

## Appendix III Research Information Infrastructure projects

Following is a summary of activities relevant to capability 5.16 that have been or are currently being conducted/supported through the EST portfolio, namely:

- Review of the Australian Partnership for Advanced Computing (APAC)
- Australian Research and Education Network (AREN)
- SII FRODO and MERRI projects
- New 2006 Systemic Infrastructure Initiative (SII) projects
- Copyright Reform
- e-Research Coordinating Committee
- PMSEIC Data for Science Working Group.

### REVIEW OF THE AUSTRALIAN PARTNERSHIP FOR ADVANCED COMPUTING (APAC)

A review of APAC forms part of the Government's normal evaluation process. As well as satisfying accountability requirements, it aims to evaluate APAC's achievements to date against agreed milestones and inform future government investments in advanced computing, data management, grid infrastructure and services. It will assist in assessing future demand for advanced computing, data management and grid infrastructure and services arising as a consequence of additional investments in research infrastructure made through NCRIS.

The Review Panel is chaired by Professor Lance Twomey, former Vice-Chancellor of Curtin University.

The other members of the Panel are:

- Dr George Collins, Chief of Research, ANSTO;
- Professor Ah Chung Tsoi, Head, Monash University e-Research Centre;
- Professor Chris Barter, former Head, Computer Science, University of Adelaide;
- Professor Michael Levine, Scientific Director, Pittsburgh Supercomputing Centre, Pittsburgh, USA;
- Dr Daniel Reed, Vice-Chancellor of Information Technology Services, the University of North Carolina at Chapel Hill, USA.

The Panel has provided a draft of its final report which will be released shortly. The Department will refer the interim report to the NCRIS Committee to inform its consideration of an appropriate level of investment of NCRIS funding in advanced computing, data management, grid infrastructure and services before it makes a recommendation to the Minister.

The Panel consulted widely, including APAC and its partners, the wider research community, and those communities related to the investment plans for the nine NCRIS capabilities currently being developed.

A summary of APAC's main activities is provided in Appendix V.

## AUSTRALIAN RESEARCH AND EDUCATION NETWORK (AREN)

The SII has provided \$88 million to date to help build the AREN, one of the largest and most advanced fibre optic research and education networks in the world.

- The AREN is operated and managed by AARNet. AARNet, the Universities, CSIRO and several state governments or agencies have invested substantially in the network.
- The AREN provides a national high bandwidth backbone with significant international links for Australian universities and the wider research community. Operating as AARNet3, the network spans the country, connecting universities and research institutes in all capital cities and many regional centres, as well as isolated research facilities such as radio telescopes. Typically, the inter-capital city backbone is 10 gigabit per second (Gbps) (except from Brisbane to Townsville where a 2.5 Gbps link exists). An increasing number of major campuses and research facilities are connected at 1 Gbps.
- The network also has a huge international footprint with high bandwidth links to Europe, Asia, Fiji, Hawaii and the US west coast, connecting to the major research and education networks of the world.
- The most recent development is the signing of contracts to establish the Mitchell link fibre optic link (Sydney-Lithgow-Bathurst-Orange-Dubbo) through a joint venture involving the NSW Government, Soul Communication, Charles Sturt University (CSU) and AARNet. In addition to providing the backbone links for the NSW Government network, it will provide dark fibre which AARNet will light to serve CSU campuses at Orange and Bathurst. The link also provides further redundancy for AARNet's NextGen network.
- There is significant potential for schools and TAFE institutes to use this publicly funded infrastructure. DEST is in discussion with schools and TAFE jurisdictions and AARNet to explore opportunities for closer collaboration between the sectors and to provide better access to educational content available across the AARNet3 and the international research and education networks it peers with.

### National status

AARNet on behalf of its members and the ARENAC negotiated with the receivers of NextGen Networks to gain access for AREN to that network. The result was an agreement to acquire access to the network for a 15-year period.

In Queensland, the Sunshine Backbone provides gigabit capacity from Townsville and Rockhampton to Brisbane. Over the next three years this network backbone will be extended to Cairns, Hervey Bay, Gladstone and the University of the Sunshine Coast.

The Tasmanian Research and Education Network (TREN) will use optic fibre provided by the Tasmanian Government on intra-state links, and the Bass Link cable currently being laid across Bass Strait (expected to be available in 2006).

The Victorian Education and Research Network (VERN) will connect sites in the greater Melbourne area and link the Bass Link cable from its landing site at Loy Yang power station in Gippsland back to the AREN (Stage 1). Stage 2 will extend the network from Traralgon into eastern Victoria as far as Lakes Entrance.

The acquisition of NextGen fibre has provided connectivity for Sydney, Canberra, Wollongong and Newcastle. The associated NextGen regional project will provide connectivity for several regional campuses, western NSW sites such as Albury, Parkes and Dubbo and north coast sites. It also connects major radio telescopes at Narrabri, Coonabarabran and Parkes.

Perth is connected via the NextGen network, however only one fibre pair is lit currently. A proposal is being developed by the WA Regional Network Organisation (WARNO) to light the second fibre pair which would allow for the deployment of dedicated network layers for particular disciplines. Kalgoorlie could be connected as part of the NextGen project. DEST and the WA universities have also co-invested in a fibre build between Perth and Mandurah which involves laying fibre during the construction of the Perth Transit Railway.

The NextGen network does not reach into the Northern Territory and negotiations of architecture and price are constrained by the absence of competition in supply. Notwithstanding these difficulties, a three-year solution is currently being negotiated with a view to providing significantly improved links to institutions and agencies in Alice Springs and Darwin. It will also provide a connection in to Charles Darwin University's intra-Territory network, thus providing better access to the resources across the AREN and the international networks it peers with for learning centres in regional and remote parts of the Territory.

SABRENet in South Australia will establish an 80km north-south high-speed metropolitan network, connecting all major university campuses, teaching hospitals, South Australian government research facilities, the CSIRO and the Defence Science and Technology Organisation in South Australia. Major routes will extend from the Adelaide CBD and reach north to Roseworthy, south to Flinders, east to Magill and west to Woodville. The Adelaide Innovation Constellation precincts of Waite, Thebarton, Mawson, Flinders and Florey, identified in the State Government's STI10 Science vision, will all be linked by the network.

#### Overseas

In November 2004, the 10 Gigabits per second (Gbps) dual link from Sydney to the west coast of the USA became operational. It was through this link that the landing data from the European Cassini-Huygens Mission to Titan collected by the Australian telescopes at Parkes and Mopra were relayed to the Joint Institute for VLBI<sup>5</sup> in Europe (JIVE) facility in the Netherlands.

AARNet now has six circuits between Australia and the US West Coast.

Dual 155 Megabits per second (Mbps) circuits:

- (i) one from Sydney to Fiji to Hawaii to Seattle, and
- (ii) one from Sydney to Seattle via an alternate path

Dual 622Mbps circuits

- (i) one from Sydney to San Jose
- (ii) one from Sydney to Los Angeles

Dual 10Gbps circuits

- (i) one from Sydney to Hawaii (Manoa) to Seattle
- (ii) one from Sydney to Hawaii (Big Island) to Los Angeles

Over the past six months, AARNet has secured 622 Mbps links to Frankfurt via Singapore in order to support further research and collaboration.

A summary of AARNet's main activities is provided in Appendix IV.

---

<sup>5</sup> Very Long Baseline Interferometry

## SII FRODO AND MERRI PROJECTS

### Managing And Integrating Large Data Sets

#### **BlueNet: The Australian Marine Science Data Network**

<u>Lead Institution:</u>	University of Tasmania
<u>Partners:</u>	University of Sydney, University of Melbourne, University of Queensland, University of Western Australia, Flinders University, University of Adelaide, James Cook University, Australian Partnership for Advanced Computing
<u>Contact:</u>	Kate Roberts (03) 6226 2873
<u>Web:</u>	<a href="http://www.blunet.org.au/index.html">http://www.blunet.org.au/index.html</a>
<u>Link with NCRIS capabilities:</u>	5.12 Integrated Marine Observing System

BlueNet will provide a highly distributed archiving facility to support the long term data curation requirements of Australia's marine science researchers. It will link vast data repositories and marine resources that currently reside in individual academic institutions and government agencies in Australia and overseas.

One third of Australia's marine R&D is undertaken by publicly funded Australian government agencies, and the remaining two thirds is carried out by state/local government and academic institutions. For Australia to maximise current and past investments in marine R&D it is imperative that, as a nation, research data are made more widely available, are more readily discoverable, and that we curate and provide access to these data in a way that more seamlessly permits its re-use. Presently, three quarters of the Australia's marine science data is contained in individual universities and not accessible to the wider researcher community.

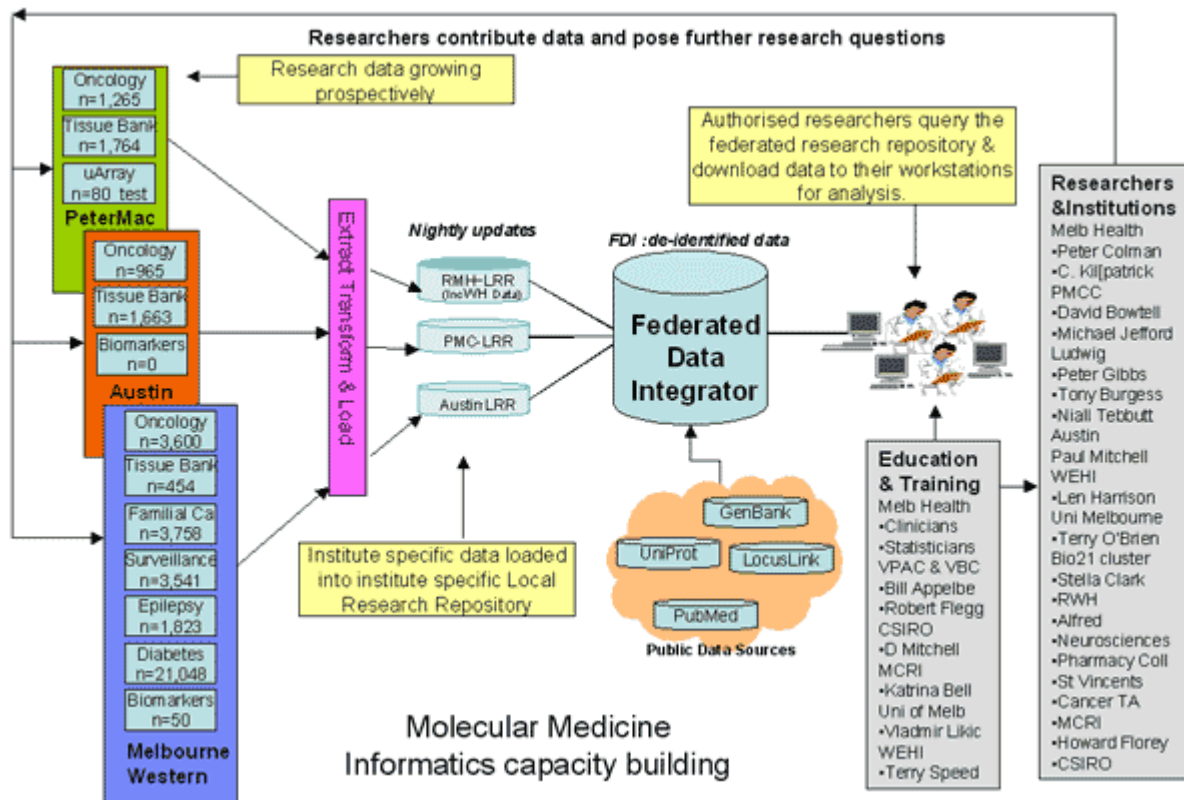
BlueNet will build infrastructure to enable the discovery, access and online integration of multi-disciplinary marine science data on a very large scale to support current and future marine science and climate change research, ecosystem management and government decision making. The objectives of the BlueNet project are to:

- Extend and enhance the existing marine science information infrastructure being developed to support the functions of the Australian Ocean Data Centre Joint Facility (AODCJF), enabling universities and other non-government agency groups to join and extend the collaborative network;
- Build a vast, on-line federated repository of marine science data that is available to support current and future marine science research, ecosystem management and government decision-making;
- Provide a distributed archiving facility to support the long-term marine data curation requirements of the university sector and Australia's marine science researchers; Demonstrably add significant value to both Australia's higher education and marine research sector.

**Molecular Medicine Informatics Model: A Multi-Institutional, Multi-disciplinary Research and Training Platform for Clinical Research (MMIM)**

Lead Institution: University of Melbourne  
Partners: Victorian Partnership for Advanced Computing (VPAC), Multiple Hospitals and Medical Institutions affiliated with the Universities of Melbourne, Monash University and the University of Tasmania.  
Contact: Marianne Hibbert (03) 9342 7066  
Web: <http://mmim.ssg.org.au/>  
Link with NCRIS capabilities: 5.7 Population health and clinical data linkage

The MMIM platform provides clinical researchers access to data from disparate existing databases across multiple disease types at multiple institutions, co-located in a virtual repository, which can be linked with publicly available research and genetic profiling data. The MMIM provides a flexible and secure method for interrogating the multiple data sources, where 80,000 records of patient data is record-linked across all databases and institutions. Researchers can extract sub-sets of data, transform where required and test hypotheses using their own analytical tools. The data is extracted nightly from all source databases where it is mapped into the Bio21: MMIM local repositories, adhering to subject area standards where appropriate. It provides a flexible way of adding new data sources, with a “plug-in” facility, and, as research requirements change with new discoveries, it has the flexibility to evolve and expand accordingly.



**TimeSync: Mapping the Global Financial System**

<u>Lead Institution:</u>	University of New South Wales
<u>Partners:</u>	ANU, Australian Stock Exchange, Macquarie University, Reuters LLC and National Electricity Market Management Company
<u>Contact:</u>	Michael Briers (02) 9236 9117
<u>Web:</u>	<a href="http://www.sirca.org.au/">http://www.sirca.org.au/</a>
<u>Link with NCRIS capabilities:</u>	N/A

TimeSync will provide real time solutions for research organisations and provide flow on benefits to financial and business organisations that need to manage and synchronise massive data transfers within secure environments. The project will develop best practice e-Research principles for data security and sustainability of time critical data using data from the Reuters financial databases in the UK, and moving this over the education and research networks to Australia and to the APAC centre at the Australian National University.

The main outcome of the project will be the establishment of a world leading and vibrant financial services research and innovation hub. Specific outcomes will include:

- Growth in size, scope, quality and impact of research, through access to a broad integrated data repository;
- Growth in number and quality of researchers using the facility, through increased ease of access whilst preserving data security;
- Growth in the quality and number of students/graduates available to the financial services sector;
- Increased reliability and integrity of the underlying databases;
- The development of robust processing systems, enabling researchers to have near to real-time access to deep history databases;
- The combination of skills developed in a range of government sponsored research initiatives which providing synergistic outcomes;
- Raising Australia's profile in financial services research and innovation;
- Establishment of a viable international university subscription model, leading to increased levels of international research collaboration; and
- Growth in research and innovation activities, and in particular providing a foundation for opportunities for commercialisation of the IP developed from use of the infrastructure.

## Technical Development and Deployment Projects

### Australian Service for Knowledge of Open Source Software (ASK-OSS)

<u>Lead Institution:</u>	Macquarie University
<u>Partners:</u>	Oxford University, Open Source Industry Association and Open Source Law
<u>Contact:</u>	Ray Warouw (02) 9850 7852
<u>Web:</u>	<a href="http://ask-oss.mq.edu.au/">http://ask-oss.mq.edu.au/</a>
<u>Link with NCRIS capabilities</u>	Relevant as a service to all capabilities

ASK-OSS provides unbiased, pragmatic advice/guidance to researchers on: selection of appropriate OSS for research; choosing appropriate OSS licenses; management/governance for OSS development; and a national service for storage and community development of OSS. Open source software is now the primary basis for most e-Research analysis and computation. The storage of complex E-Research raw data is no longer the only challenge; analysis and compute software is of equal or greater importance. ASK-OSS provides a national focal point for advice on open source software for research effectiveness as computation and analysis, particularly to support e-research, become more complex.

The outcomes of the ASK-OSS project include:

- Increasing the understanding of OSS issues in relevant Government organisations, and within appropriate sections of the research community
- Building capacity in the development, management and governance of OSS for research, and by extension to other areas of OSS adoption (such as the wider education sector)
- Assistance through education the reduction in “missteps” related to open source software, particularly in selection of appropriate open source licenses.

**Middleware Action Plan and Strategy (MAPS)**

<u>Lead Institution:</u>	University of Queensland
<u>Partners:</u>	Macquarie University, Monash University, ANU, CAUDIT, CAUL, AARNet, GrangeNet
<u>Contact:</u>	Nick Tate (07) 3365 3521
<u>Web:</u>	<a href="http://middleware.edu.au/">http://middleware.edu.au/</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

The MAPS project will identify the software and services (middleware) that are currently being used in Australia to link applications across a range of resources on networks and computer systems in Australian universities. 'Middleware' is a set of common software and services designed to connect people to a range of resources in distributed environments. These resources include major research facilities, information and communications technology infrastructure, data repositories, and other resources that support e-research, scholarly communication, and e-learning. The MAPS project will identify areas of activity in the university and research sectors, and use these results to tap into the expertise across the sector to build a strategic plan of activities and projects for an Australian collaborative middleware strategy. This project will enable other projects to leverage off common infrastructure and focus on providing new services that can be shared across the education and research sectors.

The MAPS project is working to:

- identify middleware activity that has so far been undertaken within the Australian research and higher education sector;
- identify international middleware activity and best practices that could benefit Australia;
- commission position papers from domain experts within a number of domains of middleware activity;
- develop a draft roadmap for middleware activities to support research and higher education in Australia;
- disseminate the draft roadmap widely and seek feedback from the sector;
- produce a completed roadmap and action plan.

**Open Access to Knowledge Law Project (OAK Law)**

<u>Lead Institution:</u>	Queensland University of Technology (QUT)
<u>Partners:</u>	NICTA, ANU, UNE, UNSW, University of Wollongong, Charles Darwin University, ODRL Initiative, Earlham College, Stanford University, Creative Commons, University of Melbourne (Intellectual Property Research Institute of Australia (IPRIA) and Centre for Media and Communications Law (CMCL)).
<u>Contact:</u>	Professor Brian Fitzgerald <a href="mailto:bf.fitzgerald@qut.edu.au">bf.fitzgerald@qut.edu.au</a>
<u>Web:</u>	<a href="http://www.oaklaw.qut.edu.au">http://www.oaklaw.qut.edu.au</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

Extraordinary technical advances in our capacity to disseminate and share the results and outputs of publicly funded research have not been matched by the required changes in legal and contractual arrangements.

The OAK Law project responds to the growing trend to share knowledge as a form of promoting innovation through collaborative research. The OAK Law project aims to clarify issues involving copyright management (particularly licensing) in an “open access” environment in which people and institutions are willing and wanting to share knowledge in a seamless manner. To this end the project will develop protocols for managing copyright issues in an open access environment and investigate provision and implementation of a rights expression language for implementing such protocols at a technical level.

The project will produce:

- Nationally and internationally applicable legal protocols based on the Creative Commons model that can be used to facilitate open access to copyright material;
- Legal solutions in the form of guidelines on best practice models for managing copyright issues in open access environments e.g. where further commercial publication is expected — E-Prints model;
- A Rights Expression Language (REL) that can be used to technologically enhance open access;
- An application of the legal protocols and solutions and REL to existing and proposed repositories.

## Technical Interoperability and Accessibility Projects

### **DART - Dataset Acquisition, Accessibility, and Annotation e-Research**

#### **Technologies**

<u>Lead Institution:</u>	Monash University
<u>Partners:</u>	James Cook University, CRC for Enterprise Distributed Systems Technology, University of Queensland
<u>Contact:</u>	Andrew Treloar (03) 990 53024
<u>Web:</u>	<a href="http://www.dart.edu.au/">http://www.dart.edu.au/</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

The DART project is working to support and enable researchers, end-users, and appropriate computer systems to manage the creation and collection of data and to gain greater access to data and documents by gathering, managing and archiving data and documents and managing their access so that researchers are more easily able to perform their work and do so at a much higher level of insight and productivity than was previously possible, and so that the Australian public has greater visibility of, and access to, publicly funded research.

The DART project will undertake a coordinated program of e-Research requirements analysis, software development, policy and guideline creation and prototyping to investigate how best to:

- collect, capture and retain large data sets and streams from a range of different sources;
- deal with the infrastructural issues of scale, sustainability and interoperability between repositories;
- support deposit into, access to, and annotation by a range of actors, to a set of digital libraries which include publications, datasets, simulations, software and dynamic knowledge representations;
- assist researchers in dealing with intellectual property issues during the research process; and
- adopt next-generation methods for research publication, dissemination and access.

**ARROW - Australian Research Repositories Online to the World**

<u>Lead Institution:</u>	Monash University
<u>Partners:</u>	Swinburne University of Technology, UNSW, National Library of Australia
<u>Contact:</u>	David Groenewegen (03) 9905 4563
<u>Web:</u>	<a href="http://arrow.edu.au/">http://arrow.edu.au/</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

The ARROW project is developing a national research resource discovery service using metadata harvested from participating institutional repositories. The ARROW Project is also exposing metadata to provide services via protocols, toolkits and software solutions to support best-practice institutional repositories comprising e-prints, digital theses and electronic publishing. The solutions developed are open-standards based and facilitate interoperability within and between participating institutions. A wide range of digital content types including images, text and sounds are being managed in these repositories.

As other institutions move to develop their institutional repositories, the ARROW project will provide advice on how to proceed and how to ensure that the material stored can be made discoverable and usable. In addition, the ARROW project is also working to create tools that will make institutional reporting for the RQF more efficient and effective. Through the growing repository users/developers community, the ARROW project is providing the means through which the community is able to learn from the experience of others and plan for emerging technical and cultural challenges and opportunities.

As part of this work, the ARROW project has built up a number of key partnerships with relevant bodies and projects across Australia and the rest of the world. Some of these relationships have already resulted in improved functionality and usefulness for the existing repositories. Others have created greater understanding and theoretical knowledge that is only now becoming ready to be implemented. Some relationships need ongoing involvement by ARROW to ensure that our future needs can be met.

**Australian Partnership for Sustainable Repositories (APSR)**

<u>Partners:</u>	National Library of Australia, University of Queensland, University of Sydney, APAC, University of Melbourne, UTS
<u>Lead Institution:</u>	Australian National University
<u>Contact:</u>	Adrian Burton (02) 6125 6659
<u>Web:</u>	<a href="http://www.apsr.edu.au">http://www.apsr.edu.au</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

The APSR Project is establishing a centre of excellence for the management of scholarly assets in digital format. It has an overall focus on the critical issues of the access continuity and the sustainability of digital collections. Building on a base of demonstrators for digital continuity and sustainability embedded in developmental repository facilities within partner institutions, the APSR project will contribute to national strength in this area by encouraging the development of skills and expertise and providing coordination throughout the sector. It will actively provide international linkages and national services.

The University of Sydney Library is working with the APSR partners in developing a set of guidelines to assist repository planning, creation and management. A variety of presentation approaches is required to allow choice and flexibility in the browsing and retrieving of the mountains of available information. Where some people might feel comfortable doing a free search of the information, others may prefer answering a series of questions to narrow down the scope or empathising with scenarios that describe real-life situations.

An example of APSR's contribution is the Sustainability Guidelines for Australian Repositories (SUGAR) which will provide an online information service for people seeking guidance on digital sustainability issues. The guidelines, drawing on international best practice, will be developed as a prototype and tested with reference to digital repository projects within and beyond the University of Sydney. The underlying data management system will be sufficiently flexible to enable presentation of content through a variety of models and scenarios. The service will include advice and link to sources on topics such as project planning, costs, digital conversion standards and processes, legal and copyright issues, metadata schemas and modes of access.

**Australian Digital Theses Project**

<u>Lead Institution:</u>	University of New South Wales
<u>Partners:</u>	CAUL, University of Melbourne, University of Queensland, University of Sydney , ANU, Curtin University of Technology, Griffith University
<u>Contact:</u>	Howard Amos (02) 9385 3965
<u>Web:</u>	<a href="http://adt.caul.edu.au/adtariic.html">http://adt.caul.edu.au/adtariic.html</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

Completed in 2006, the Australian Digital Theses (ADT) project has established a distributed database of digital versions of theses produced by the postgraduate research students at Australian universities. With the inclusion of metadata tags in the documents which are given a higher weighting by the more sophisticated search engines, 150,000 theses are available worldwide via the web; 6000 of which are full-text available. This number is growing rapidly as more researchers are depositing their works in accessible repositories. In addition, several institutions in New Zealand have also joined the project. The ADT program is designed to improve access to, and enhance transfer of, the research information contained in theses by providing a full text version available from the desktop via the web. The ADT also provides a new model for deposit and archiving of theses that takes into account the tools and technologies that students are now using to prepare their theses.

The program has two major components, digitisation of theses as part of the deposit process and the digitisation of a selected number of frequently requested existing theses. As each University is responsible for maintaining an archival copy of the theses of their own institution, each participant in the program will mount their own theses on a server located in their respective institution. The participants will use the same database configuration, standards and metadata system to ensure compatibility. The document format will be Adobe Acrobat Portable Document Format (PDF) ensuring that the data is independent of the platform on which it is created. Adobe PDF ensures that a high quality printed version can be provided if needed. Acrobat is relatively easy to use, with a high quality free reader readily available. PDF has also become an electronic publishing standard.

**E – Security Framework for Research**

<u>Lead Institution:</u>	University of Queensland
<u>Partners:</u>	Macquarie University, CAUDIT, APAC, AARNet
<u>Contact:</u>	Nick Tate (07) 3365 3521
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

This project is establishing an E-Security framework which integrates different two types of security systems, Public Key Infrastructure (PKI) and Shibboleth, to foster collaboration and enable the secure sharing of resources and research infrastructure within Australia and with international partners. The project leverages off existing work in both areas, building on the advantages of these different systems and creating a platform to enable the secure sharing of resources and research infrastructure.

Secure access and the authentication and authorisation of researchers, who access services and infrastructure across global networks, are fundamental building blocks for e-research. The E-Security framework is developing solutions that address these issues as well as a system where a researcher can get information from another university, without having to use different security measures to access it.

The E-Security Framework has four main objectives:

- Build upon the existing Public Key Infrastructure (PKI) standards project and move PKI into production for the Higher Education and Research Sector.
- Build upon the existing PKI and MAMS projects and the Production PKI project to develop models and pilot implementations of a common trust federation which would support both PKI and Shibboleth and therefore support a common approach to authentication and authorisation across the sector.
- Reduce the barriers for entry to PKI for all universities and research institutions by providing cost effective access to a free or low cost Certificate Management System for the sector (including access to the source code).
- Investigate the requirements and develop appropriate technologies to allow the APAC Grid infrastructure to become properly Shibboleth aware. It will provide opportunities for research activities in high-performance computing and large-scale data initiatives to test the functionality and scalability of the Shibboleth authentication architecture and associated authorisation architectures being developed by groups such as PERMIS.

**Regional Universities Building Research Infrastructure Collaboratively (RUBRIC)**

<u>Lead Institution:</u>	University of Southern Queensland (USQ)
<u>Partners:</u>	UNE, University of the Sunshine Coast, University of Newcastle, Massey University (NZ). Project is a joint initiative between DEST and the UK JISC (Joint Information Systems Committee).
<u>Contact:</u>	Associate Professor Alan Smith (07) 4631 2296
<u>Web:</u>	<a href="http://rubric.edu.au">http://rubric.edu.au</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

Regional universities often face different challenges to larger institutions in developing information management infrastructure. At the same time, these smaller universities may have some advantages over larger institutions through their greater agility for adopting innovations with less bureaucracy. The RUBRIC partners recognize the need to keep pace with larger institutions in building appropriate infrastructure to support research and allied educational processes. Repository infrastructure becomes a core mechanism to support a University's research mission. It brings a responsibility to manage levels of access management, a scrutiny of content retained for the purpose of establishing collection development policies, retrieval, re-use and federation of objects and to explore links between repository content and full text sources. For smaller organizations these are significant areas of specialty where a collaborative approach is of considerable benefit, if not strategically vital.

The RUBRIC Project aims to meet the needs of smaller and regional universities by developing sustainable infrastructure for the deployment of best practice emerging from the SII first round projects known as FRODO (Federated Repositories of Digital Objects) and other projects funded in the second round known as MERRI (Managed Environments for Research Repository Infrastructure).

The project aims to achieve the following outcomes for its partners:

- Adopt "best practice" emerging from the FRODO projects to be able to contribute regional experiences to debate in the Australian context;
- Provide a centralized base for assistance to RUBRIC partners, which are generally smaller regional organizations with limited resources;
- Set up a first generation solution for all partners (or second generation for partners with some existing experience, such as USQ);
- To build a central resource for investigating, researching and making recommendations on IR developments - where issues can't be resolved, they will at least be documented and thrown open for debate;
- Building a knowledge base for software evaluation, data management issues, community engagement and monitoring of national and international developments in this field.

**Meta Access Management System Project (MAMS)**

<u>Lead Institution:</u>	Macquarie University
<u>Partners:</u>	National Library of Australia, education.au, Telstra Research Labs, University of Southern Queensland, UNE, University of Tasmania, University of Newcastle – Australia, UWA, Curtin University of Technology, Edith Cowan University, Murdoch University, The University of Notre Dame – Australia, Internet2/MACE 'Shibboleth', University of Ulster, National Library of New Zealand, Massachusetts Institute of Technology, Open Knowledge Initiative (OKI), Joint Information Systems Committee (JISC), Centre for Educational Technology Interoperability Standards (CETIS) UK
<u>Contact:</u>	James Dalziel (02) 9850 7539
<u>Web:</u>	<a href="http://www.melcoe.mq.edu.au/projects/MAMS/">http://www.melcoe.mq.edu.au/projects/MAMS/</a>
<u>Link with NCRIS capabilities</u>	Relevant to all capabilities

The MAMS project is developing next generation access and identity management infrastructure and related common technical services to support research effectiveness, and implementation advice and programs, together with support for practical implementations among project partners.

The MAMS project is developing a prototype middleware system for access and identity management, supported by additional common services such as search services, digital rights management and metadata management. This will include the implementation of an “umbrella” or meta access management system for intra-institutional authentication and authorization based on a multi-modal approach to the process of single-sign-on that recognizes the realities of current legacy environments. In addition, implementation of an inter-institutional authentication and authorization regime based on attribute exchange and automated decision-making using machine-readable policy

The MAMS project is working to disseminate its accomplishments with a focus on development and trials of implementation of new approaches to access and identity management and other common technical services. Given the focus of this project on middleware, this project provides core information infrastructure to ensure that the national investment in hardware and networks is complemented by the necessary software infrastructure to unleash the potential of higher education research information infrastructure.

## NEW 2006 SYSTEMIC INFRASTRUCTURE INITIATIVE (SII) PROJECTS

The following are newly funded projects (July 2006) structured as next stage activities for some of the existing SII projects described previously. These projects will reinforce the momentum already established by:

- Progressing information infrastructure technologies into a new phase of deployable production-ready software and services.
- Enlarging the network of digital repositories and the community of users through direct collaboration with phase 1 NCRIS capabilities and other institutions in the sector.
- Supporting universities and other research agencies in their ongoing use and development of digital repositories, allowing greater discovery, preservation, access, and sharing of research data and research outputs.
- Facilitating the introduction of the emerging RQF through development, integration and deployment of technical solutions designed to support researchers and research institutions in meeting RQF reporting requirements.
- Further developing the underpinning technologies supporting authorisation, authentication, and access including identify management solutions.
- Development of tools to enable researchers to publish online in the current regulatory environment.

### **Australian Research Enabling Environment (ARCHER))**

Lead Institution: Monash University  
Link with NCRIS capabilities Relevant to all capabilities

This project will build on the architecture and prototype software developed by the Dataset Acquisition Accessibility & Annotation e-Research Technologies (DART) project to adopt a common information management architecture and infrastructure across the many data intensive research areas represented in the 9 high priority capability areas under NCRIS. It will similarly address the information management needs of capability areas in the Social, Behavioural and Economic Sciences and in the Humanities and Creative Arts.

### **Research Activityflow and Middleware Priorities (RAMP)**

Lead Institution: Macquarie University  
Link with NCRIS capabilities Relevant to all capabilities

The RAMP project aims to improve national research effectiveness by addressing two challenging aspects of a national research information infrastructure: the development and implementation of open standards authorisation for protected repositories; and research into and demonstration of people-oriented research workflows (often referred to as research activityflows).

### **ARROW-2**

Lead Institution: Monash University  
Link with NCRIS capabilities Relevant to all capabilities

The Australian Research Repositories Online to the World (ARROW) project has been very successful in providing tools to enable accessibility and discoverability of research from institutional repositories. ARROW Stage-2 will build upon this success to support the building of institutional repositories for project partners and advise new members on how to proceed and what to expect. An important aspect of

the ARROW Stage 2 project will be the establishment of sustainable pathways for institutional repositories beyond the term of the project.

**ICE-RS**

Lead Institution: University of Southern Queensland  
Link with NCRIS capabilities Relevant to all capabilities

ICE-RS will create open standards based technical solutions to facilitate and encourage the efficient creation of flexible documents in the process of conducting and reporting on research. ICE-RS will deliver a research authoring environment that assists researchers to systematically create, structure, and manage their publications and reports, and aids the automation of research workflows. The project will build on existing work undertaken at the University of South Queensland, and The Australian National University along with contributions from the Regional Universities Building Research Infrastructure Collaboratively (RUBRIC) project.

**Legal Frameworks for e-Research**

Lead Institution: Queensland University of Technology  
Link with NCRIS capabilities Relevant to all capabilities

This project will extend and reinforce the work already being undertaken by the Legal Protocols for Copyright Management for Open Access project. It will create an online intellectual property and licensing database that will augment and interoperate with the JISC funded Securing a Hybrid Environment for Research Preservation and Access (SHERPA) project. One part of the project will examine and categorise existing publishing agreements of key Australian and other relevant publishers and present these results via a web interface. The second part of the project will map out a sophisticated legal framework for e-Research and collaborative innovation.

**APSR-2**

Lead Institution: ANU  
Link with NCRIS capabilities Relevant to all capabilities

APSR is an open partnership of research and higher education institutions, funded under SII, committed to strengthening the national research infrastructure through the development of digital repositories and the provision of associated research-linked discovery, access and management services. This extension to APSR will build on the existing strengths and success of the partnership and continue its role as a key organisation for promoting best practice and expertise in managing digital collections.

## **COPYRIGHT REFORM**

DEST, through a number of SII-funded projects and with specific internal DEST funding, has advocated the education and research sectors' interests in copyright law reform for several years.

In particular, the amendments to the Copyright Act 1968 arising from the Australia-United States Free Trade Agreement (AUSFTA) shifted the copyright balance in favour of copyright owners (extension to copyright term, anti-circumvention provisions for Technological Protection Measures amongst others). Recent review processes have sought both to finalise implementation of those provisions and to re-examine copyright exceptions in light of technological change and common consumer practice.

DEST has played a central role in coordinating the responses from a range of education sector interests to recent copyright law reform processes. DEST made substantial submissions to the Attorney-General's Department's inquiry into Fair Use and Fair Dealing Exceptions in the Digital Age, the House of Representatives Legal and Constitutional Affairs Committee's Inquiry into Technological Protection Measures Exceptions and the Attorney-General's review of Carriage Service Providers.

DEST's stakeholders are both creators and users of copyright. DEST has an interest in maximizing the commercialisation and national return on research through the collaboration of industry, universities and CSIRO.

There is also an important public interest in encouraging innovation, and in ensuring effective access to copyright materials for education and research purposes. Such access determines the opportunities for students, teachers and researchers to take their place in a competitive global information economy.

As well as seeking an appropriate balance between different interests, DEST's approach to copyright law reform has been that changes to the Act should seek:

- to maintain technological neutrality between the operation of copyright law in the digital and non-digital environment;
- to maintain neutrality and consistency of treatment as between different categories of subject matter;
- to balance flexibility and certainty;
- to maximise the simplicity of the provisions; and
- to lower compliance costs for creators and users.

## E-RESEARCH COORDINATING COMMITTEE

In September 2004, in *Information Technology: Connecting an Innovative Australia*, the Government made an election commitment to implement a coordinated structure for e-Research, modelled on the UK e-Science Programme.

The Government's election commitment is being carried out in two phases:

- An initial investigation and consultation phase, "to involve relevant stakeholders in an overarching e-Research Coordinating Committee"; and
- An implementation phase, to put into place the Committee's recommendations, subject to Ministerial approval.

In April 2005, the e-Research Coordinating Committee was established jointly by the Minister for Education, Science and Technology and the Minister for Communications, Information Technology and the Arts to provide expert advice to the Government about developing Australia's e-Research capacity. The Ministers appointed Dr Mike Sargent AM to chair this Committee.

To provide it with expert advice in key areas, the Committee established a Technical Working Group and a Computer Sciences Expert Group. It also sought advice from the participants in relevant programmes funded under the Systemic Infrastructure Initiative.

In December 2005, The Ministers accepted the Interim Report of the Committee: *An e-Research Strategic Framework*, as the basis for further work to develop an e-Research implementation strategy. The Committee has now completed its Final Report: *An Australian e-Research Strategy and Implementation Framework* following extensive consultations on the e-Research agenda with stakeholders around Australia, in parallel with the NCRIS Committee consultations.

Feedback from stakeholders and experts engaged by the Committee confirms strong support from state governments, universities, industry stakeholders and the research community, for a national strategic framework to coordinate and accelerate the development of Australia's e-Research capabilities. This is recognised as a priority if Australian research is to remain relevant and competitive and we are to avoid creating incompatible research silos.

Key findings of the Committee are:

- Skills acquisition and skills transfer are vital in developing Australia's e-Research capability.
- An e-Research Centre, consisting of a set of nodes located in areas of research concentration, should be established to assist the research sector to develop e-Research capabilities by coordinating outreach and awareness raising, promoting skills transfer, providing support and undertaking strategic research.
- An e-Research Committee, chaired by a high profile e-Research Leader / champion, should be appointed to oversee the implementation of the e-Research agenda. (The experience of the UK e-Science initiative strongly suggests that the appointment of a well respected member of the research community, recognised by government, the research sector and industry as the visible face of e-Research is essential in engaging the research community in e-Research.)
- The Government should establish a working group to develop an Australian Research Data Strategy that will support a standardised national approach to the management of data collected, generated and used by the Australian research community.

The e-Research Strategic Framework outlined in the Committee's Final Report is designed to be implemented over a period of five years. It is expected that the outcomes of implementing this Framework will be:

- Australian researchers will be world leaders in the use of advanced and innovative information and communications technologies, to achieve internationally recognised, high quality research outcomes across Australia's national research priorities;

- Australian researchers will be able to access data, instruments, computing capability and to collaborate with each other, through advanced ICT, enabling them to engage readily in collaborative research and contribute significantly to the solution of major national and international research challenges;
- Australian researchers will have the necessary education, training and skills, and support from ICT and information management specialists, to use advanced ICT efficiently and effectively;
- The implementation of e-Research capabilities across the Australian research sector will provide a leading influence on the uptake and enhancement of such technologies by Australian business and industry; and
- The Australian community and economy will benefit from the advanced capability enabled by e-Research.

The activities of the e-Research Strategic Framework are intended to be integrated closely with initiatives under NCRIS and the Systemic Infrastructure Initiative, in order to inform and maximise returns on these key infrastructure investments. The Framework will also build upon initiatives in the Communications, Information Technology and the Arts portfolio, including the Advanced Networks Programme, Connect Australia and the National Broadband Blueprint.

The e-Research capabilities to be developed under the Framework are expected to underpin future Government initiatives in e-Health, e-Education and e-Government.

## **PMSEIC DATA FOR SCIENCE WORKING GROUP**

The Prime Minister's Science, Engineering and Innovation Council working group on Data for Science was essentially established to address the central question of "Should Australia have a national approach on managing data for science?"

The Working Group's Terms of Reference are below.

The PMSEIC Working Group will:

1. Provide an overview of current approaches to the management of large amounts of scientific information and data for research;
2. Outline the issues surrounding the nature of data and data repositories and libraries, including access, storage, authentication, sustainability, protection and standards for interoperability;
3. Examine whether there are advantages in Australia for a single virtual repository or key multiple domain-specific repositories of scientific information covering all research institutions, both Universities and Government (and, if they wanted to, private companies);
4. Identify issues relating to the development of an Australian virtual repository and infrastructure mediating access to the repository;
5. Identify issues relating to industry participation in the development of, and access to, an Australian virtual repository or repositories;
6. Identify strategies that could be introduced to improve access to research undertaken by publicly funded research agencies (eg. protocols around repositories), including international access and collaboration;
7. Identify a data management strategy to ensure Australia's scientific sector is globally competitive and provides benefits to the Australian economy, environment, and society;
8. Take into account the conclusions and recommendations of the e-Research Coordinating Committee.

The PMSEIC Data for Science Working Group was established during 2006 to examine and advise on directions for managing the vast amounts of data that currently exist as a result of the flood of data being generated from scientific research, observational projects, instruments, national and international collaborations, data mining and analysis.

It is expected to deliver a report and present its findings to the PMSEIC meeting to be held in December. The working group has sought input from outside of its membership and consulted with experts from academia, research agencies, other government bodies, and from industry. The e-Research Coordinating Committee's report has been provided as input to the working group, as has information on the Systemic Infrastructure Initiative funded projects.

## Appendix IV AARNet

### Statement of Capabilities/Constraints for the Research Community

This document has been prepared to give NCRIS Facilitators, in particular the Facilitator for Platforms for Collaboration, a simple understanding of AARNet's current and prospective capabilities to provide network services to the Australian research community. It also attempts to explain some of the relevant constraints under which AARNet operates.

AARNet is a not-for-profit company owned by its members, comprising 38 of Australia's 39 universities and the CSIRO. These members elect a Board of Directors which determines AARNet's strategy, policies and finances.

The systems and services provided by AARNet include the following:

Basic network:

- National backbone
- International circuits and connections
- Commodity Internet and Research & Education traffic

Network operations (NOC)

Network consulting services (eg design of extensions, customer connections)

Additional network services:

- Basic network services (eg IPv4, IPv6, QoS, Multicast)
- Overlay networks (eg configurable light-wave circuits)
- Content mirror
- Peering with other networks & providers

Value-added services:

- VoIP
- Video over IP
- ENUM
- Middleware
- eduroam
- Network-connected filestore

The following analysis provides detail of the elements of these services likely to be of greatest interest to the research sector, especially as represented by the NCRIS process (see [http://www.dest.gov.au/sectors/research\\_sector/policies\\_issues\\_reviews/key\\_issues/ncris/](http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/ncris/)).

### 1. Executive Summary

The principal service of AARNet likely to be of interest to the researcher is its basic high-capacity network, providing connectivity between points of interest, whether nationally or internationally. In addition to this network connectivity and capacity, AARNet provides a range of related network services, as well as value-added services and applications, that are likely to be of interest to the research community.

AARNet's network and charging structure has been designed specifically to eliminate all direct traffic volume charges for transfers between members of the international research and education community. This means that, provided the end points of any transfer are within that community, there will be no charge from AARNet for that transfer, regardless of its size or length of time to deliver it.

There are, however, certain caveats to this freedom from charges, which are detailed in the rest of this document:

- a) the hosting institutions for each end of a communications session must be connected to the global research & education network; this may require the establishment of physical connections and completion of access agreements;
- b) users must verify that the end-points of a transfer are indeed within the designated research & education community (see 2(g) below);
- c) users also need to verify that, where their traffic flows over sections of the network that are not under AARNet's control, the conditions applying are consistent with those which AARNet applies;
- d) end users should also verify in particular that they do not incur any charges for their traffic from their host institution (AARNet very strongly encourages its members not to impose any such charges), nor from any intervening non-AARNet network segment operator;
- e) certain services require specific network characteristics to be enabled across the full extent of the network, to achieve full end-to-end capability; these functions may require these other agencies to enable them over their sections of the network (see 2(h) below).

## 2. Basic Network

The network operated by AARNet comprises a number of different components, as follows:

- a) **Backbone:** A substantial backbone based on fibres acquired on a 15-year IRU from NextGen (which is also contracted to maintain the infrastructure to support that network). These run from Brisbane via 2 geographically distinct paths to Melbourne via Sydney and Canberra. They extend from Melbourne via Adelaide to Perth along a single geographic path with diversity achieved by taking lower capacity services from Optus..
- b) **Optical Circuits or Overlay Networks:** Between Brisbane and Adelaide, the fibres are lit using AARNet's own DWDM equipment, enabling many centres along those paths to be connected at high speed, and for optically separate "overlay" networks to be configured for customers with special needs (eg the several campuses of a regional university, the various telescopes in inland NSW, the partners in APAC). Currently, the Adelaide-Perth circuit is a 10Gbps managed circuit provided by NextGen; however, it is hoped to upgrade this also in the near future to DWDM equipment (the so-called Adelaide-Perth 2<sup>nd</sup> Fibre Project).

Current connections are typically 1Gbps. 10Gbps is possible with existing equipment, but requiring the purchase of special interface cards. By upgrading the DWDM equipment it would be possible with the current fibres to provide circuits up to an aggregate of 320Gbps (though this option is currently expensive).

The reach of this network can be seen in Diagram A, attached.

The use of the AARNet Network is governed by the AARNet Access Agreement or Regional Connect Agreement as the case may be which reflects among other things the terms of the various supply agreements, in particular NextGen. Simply put, the AARNet Network can be used for any research or educational purpose but, for instance, may not be used by third parties for commercial purposes. Use of any AARNet network or value-added service is covered by AARNet's Acceptable Use Policy (see [http://www.aarnet.edu.au/publications/access\\_policy\\_2005.html](http://www.aarnet.edu.au/publications/access_policy_2005.html)). Most services, including the optical circuits, are available only to AARNet members and associate and affiliate members (see <http://www.aarnet.edu.au/services/nonmembers.html> for a description of what organisations are eligible to become members).

- c) **Extensions to Darwin & Tasmania:** the connections to Darwin and Tasmania are different historically due to distance and lower traffic volumes. These currently are managed services supplied by Telstra, and provide considerably slower speeds. In the next 2 years, it is planned to upgrade the

circuit to Tasmania by leased dark fibre (as for the basic backbone network). In the longer term it is hoped to be able to find a way to do the same for Darwin.

- d) **Compatible Extensions:** the above basic network has been extended by the acquisition, in conjunction with certain groups of customers and other network service providers, and often with special SII grant funding, to reach various sites not covered by the basic NextGen fibre network. These include a 2.5Gbps extension from Brisbane north to Townsville, and thence (by slower circuits) to Cairns. Also in various parts of regional NSW; and also from Perth to Mandurah.

In each case, the nature of the circuits, the supplier, the duration of the arrangement, and the terms under which it can be used, differ; this is as dictated by geography, the available providers and suitable technology. The goal in all cases has been to achieve at least 1Gbps capability, but this has not always been achieved.

Generally, the network extensions are negotiated and managed by AARNet (usually through the local network service provider), and under terms that are consistent with AARNet's contracting principles for an Australia-wide network (AREN) such that it can supply an end-to-end service under the relevant Agreement

- e) **Other Extensions:** In a few other cases, extensions to the network, some with substantial capacity, have been undertaken by regionally-based groups (eg VERN in Victoria and SabreNet in Adelaide), often with financial assistance from DEST's SII or other government sources. These network extensions are typically managed independently of AARNet, and the terms and conditions governing use may differ from AARNet's, in some cases being more restrictive. The effect of this may be to limit the full end-to-end delivery of certain AARNet services, or for certain customers
- f) **International Connections:** as depicted in Diagram B (attached), AARNet has extensive international connections and capabilities. Most of these are for general or commodity Internet traffic (ie any kind of traffic), while others are limited to carrying R&E traffic as determined by the supplier (for instance where capacity has been made available under generous terms, such as by SXTRANSPORT across the Pacific). AARNet routes the traffic according to the destination of the traffic and in accordance with the policies set up by the contract.
- g) **On-Net and Off-Net:** AARNet connects in various ways to the commodity Internet (eg by peering at some Australian Internet Exchanges, and by buying capacity at other Internet Exchanges, locally and internationally). Traffic over these connections is termed "off-net", and is subject to traffic charges, which are passed on to customers on a volume-linked price basis. This off-net traffic currently comprises about 85% of AARNet's total traffic.

AARNet also connects directly with (ie peers with, free of traffic charge) key international R&E networks in the USA and Europe (eg Abilene, the Internet2 network, and GEANT2 the pan-European R&E network), which in turn connect with most R&E networks around the world. Traffic over these networks (as well as over AARNet itself between members) incurs no volume-based charge. These arrangements have been put in place specifically to support high-volume research endeavours, where a volume-based price would be a significant deterrent; these links are currently only lightly used. It is possible (and highly advisable) to determine in advance if a particular proposed traffic flow will be on-net or off-net (see <http://www.aarnet.edu.au/cgi-bin/traffic.cgi>), providing researchers with the confidence that their data will be covered by their on-net subscription charges.

All members (as well as associates) of AARNet are entitled to this subscription-based on-net arrangement having no traffic volume charges. If a member wishes to enter a research partnership with a non-member, under which significant data transfers may take place, then it may be advisable to arrange for that second party to become an associate of AARNet. This is normally possible for research purposes, but a physical connection and membership fees would be involved. Such arrangements are treated on a case by case basis.

- h) **End-to-end Network Characteristics:** certain services, such as Access Grid and Compute Grid, require specific network characteristics (eg Multicast) to be enabled; although these functions are available across the AARNet network, they may not be enabled over sections which are not under AARNet's control, such as within member institutions; institutional firewalls may also represent

impediments to unfettered access; users should verify that these anomalies are not present – AARNet can advise and assist in ensuring there are no such impediments.

Although all of the above comprises the so-called AREN (Australian Research & Education Network), it is clear that it is not homogeneous in a variety of ways. AARNet as the designated operator of the AREN must do so in accordance with the various different arrangements and conditions applying in each case.

### 3. Network Operations

AARNet operates a Network Operation Centre (NOC – see <http://noc.aarnet.net.au/>), which monitors network performance and collects data on a wide variety of network activity. The management and collection of this data is again limited to those links that AARNet directly owns and manages. This means that some end-to-end connections may not be as “visible” to the NOC as for the bulk of the network. Much of this data collection and error detection is automated, employing software called Nagios (see <http://www.aarnet.edu.au/network/>). A 24x7 service is provided, with staff available for call-out after hours – these calls can originate from automated systems sending SMS to on-call staff, or via the outsourced after-hours call centre (so customers can place calls at any time) – see <http://www.aarnet.edu.au/contacts/nochelp.html>.

### 4. Network Consulting Services

AARNet constantly undertakes consulting assignments for customers and others in the design of network extensions, overlays and other basic network services, drawing on its wealth of experience in dealing with tier 1 fibre suppliers, tier 2 network service providers, network equipment and software suppliers, typically involving leading edge equipment and services. Many network configurations and services (eg configured light-waves) are not yet available from commercial suppliers, and AARNet undertakes its own designs and configurations, drawing on its contacts with similar advanced research network operators (such as Internet2) overseas.

### 5. Additional Network Services

The network provided by AARNet currently supports a range of advanced services, which are integral to the network infrastructure and which extend over all (or nearly all) of the network. They are typically not charged for in addition to basic network charges.

These services include IPv6, Multicast and QoS, which are required by certain advanced applications such as Access Grid. It should, however, be noted that while these services are activated on the backbone, there may be issues within a connected organisation or campus that prevent full end-to-end compliance.

AARNet technical staff can advise on what the requirements may be for any particular application, and can assist with appropriate configurations and implementation.

AARNet’s network, as indicated above, is capable of supporting dedicated circuits via independent wavelengths (termed “overlay networks”), over most but not all of its length. These are usually established as 1Gbps circuits between specific end points (two or more). Once again, staff are able to advise on, design and set up such circuits to suit specific requirements. In due course, it should be possible to establish user-configurable circuits of this kind (User-Controlled Light Paths, UCLP), which can be established by researchers themselves (or with the assistance of their own technical staff) on an *ad hoc* basis or for short-lived projects. This service incurs additional costs, so cannot be provided (at this stage) free of charge.

An integral part of the network is a Content Mirror, which supports local, on-net access to a wide range of popular software and other files, so providing faster and no cost to these files; see <http://mirror.aarnet.edu.au/>.

In a similar vein, aimed at reducing the cost of traffic, AARNet peers its network with a range of other networks in Australia and overseas. Peering is usually free of charge (excluding cost of equipment), and provides un-tariffed bi-directional traffic flow between the two or more peered networks. Of particular significance are the peers with Internet2, GEANT2 and other R&E networks worldwide. AARNet is also working to gain on-net access to a range of content appropriate to the Australian higher education and research sector, including such content as Australia Screenonline, ResearchChannel, the ABC. Researchers generally do not need to be aware of these peering arrangements (other than to note which are available on an un-tariffed basis).

## 6. Value-Added Services

AARNet also is providing a widening range of value-added services and applications which are layered over the network which may be relevant to researchers. Typically, these services incur additional charges, though in some cases (eg VoIP) are low. Most of these services are currently being reviewed with a view to expanding them and making them easier to use.

The services currently provided are detailed at <http://www.aarnet.edu.au/services/> and include:

- a. Voice-over-IP (VoIP) – AARNet provides a VoIP service over its network so that, for example, interstate phone calls can be made between participating members at negligible cost (see <http://www.aarnet.edu.au/services/voip/index.html>);
- b. Video and Audio over IP – AARNet provides a general multi-point audio/videoconferencing service, based on its recently-upgraded MCU, which is currently being enhanced and made simpler to use, incorporating a booking system, integration with desktop calendaring systems like Outlook, and other collaboration and presence tools; plans are being made to add a session recording service, and to experiment with immersive multimedia solutions (see <http://cbr-b-voip1.aarnet.net.au/cgi-bin/bookmcu>);
- c. ENUM – AARNet is one of a 2-3 accredited Registrars in Australia for the ENUM communication medium numbering system, set to rival IP addresses – see <http://www.aarnet.edu.au/services/enum/index.html>;
- d. eduroam – AARNet provides the Australian and APAN regional nodes for this global wireless roaming system, and provides support and promotion; it is based on authentication at a researcher's home institution by means of a Radius system and the 802.1X protocol; once set up at the home institution, and a certificate has been loaded onto the user's laptop, it automatically connects to the Internet at any other institution worldwide which has also been set up (see <http://www.eduroam.edu.au/>).
- e. Network-connected filestore (RETAIN) – as well as hosting the above Content Mirror service, AARNet's 15TByte file store can also be used by researchers for holding file on a temporary or permanent basis (see <http://mirror.aarnet.edu.au/>); AARNet also has taken responsibility for the former GrangeNet filestore.
- f. Middleware services – AARNet provides several middleware services, including hosting the WAYF service for the MAMS Test Shibboleth Federation, maintaining a library of information about middleware initiatives locally and internationally, and co-chairing the APAN middleware stream (see <http://www.aarnet.edu.au/engineering/middleware/index.html>); these services are currently being reviewed and expanded.

## Appendix V APAC, the National Facility, the National Grid<sup>6</sup>

APAC was established to strengthen Australia's advanced computing capabilities by installing a peak computing system at the APAC National Facility and by increasing the national expertise to support users of advanced computing systems.

Over the last three years the partnership has grown in expenditure from \$33M pa to \$45M pa, by research users from around 1200 to nearly 2000 and in staff either employed or supported from around 90 to 168.

The context for the second stage of APAC (2004-06) was influenced by the increasing demands for advanced computing services in the Australian research community and by the major initiatives related to 'cyber-infrastructure' in many other countries.

These demands were being addressed by major government initiatives in many other countries such as the USA, nations of the European Union (especially the UK), Japan, Korea, Taiwan and Singapore. The initiatives were focussed on the provision of advanced computing facilities, grid services and broadband communications to support the emerging eResearch 'revolution'.

The need for Australia to participate in this global revolution guided APAC to develop a mission around seamless access to its facilities as part of the global research infrastructure, which was expressed as:

*'Providing the national advanced computing and grid infrastructure for eResearch'*

APAC therefore supports researchers by providing a peak computing service based on the National Facility and grid services that enable seamless access to distributed computing and data management facilities through the National Grid.

The 'open' architecture for the National Grid enables access to institutional resources, on-line instruments, sensors and other data sources and enables interoperability with grids in institutions and other countries. The impact of this architecture is being seen in recent discussions with organisations on incorporating their facilities into the National Grid.

The major sources of funding into these initiatives over the last three years is indicated to the right, noting the difference in the Total Cash Income and Expenditure represents a cash carryover of around \$2m at the end of 2006.

<b>TOTAL for 2004-2006</b>	<b>CASH</b>	<b>IN-KIND</b>	<b>TOTAL</b>
	(\$'000)	(\$'000)	(\$'000)
<b>INCOME</b>			
<b>SII</b>	29,000		29,000
<b>Partners</b>	9,000	24,730	33,730
<b>Other</b>	4,200		4,200
<b>TOTAL</b>	42,200	24,730	66,930
<b>EXPENDITURE</b>			
<b>National Facility Program</b>	19,810	6,975	26,785
<b>CT&amp;T Program</b>	3,320	3,150	6,490
<b>Grid Program</b>	11,800	12,400	24,200
<b>EOT Program</b>	3,000	2,205	5,205
<b>Management</b>	2,170		2,170
<b>TOTAL</b>	40,120	24,730	64,850

While the remainder of this summary focuses on the National Facility and the National Grid, APAC has also operated a Computational Tools and Techniques Program (the CT&T line in the above table) and an Education, Outreach and Training program (the EOT line in the above table).

### National Facility

The National Facility operates on a resource share model whereby APAC, its partners and other organisations are able to obtain resources on the National Facility in proportion to their cash and in-kind contributions. Under the resource share model, APAC's funding to the National Facility purchases high end compute resources available to all Australian researchers through a Merit Allocation Scheme.

The National Facility is also supported by 15.5 effective full time staff. The roles of the staff are to manage the facility, to help users optimise codes, organise software and the user environment, perform system development and administration, develop training courses, and to service the granting committees.

<sup>6</sup> These notes are extracted from the extensive documentation provided by APAC to the APAC review panel.

The current major system provided by the APAC National Facility is an SGI Altix comprising:

- 1928 1.6Ghz Itanium2 processors (1920 dedicated to batch jobs) with 5.6Tbytes main memory
- SGI's NUMalink4 interconnect providing 3.2 Gbytes/s bandwidth per link and <2us MPI latency
- 30 Tbytes of global file system with an aggregate bandwidth of > 2Gbytes/s
- A total of over 70 Tbytes of scratch disk local to the compute nodes

In addition, the National Facility operates a small 156 processor cluster fitted with 1.4Tbytes of shared file space and 15 Tbytes of scratch space, intended to take small jobs off the main system. The load shift has been highly successful and an upgrade to a larger cluster is foreshadowed.

The National Facility also provides a large Data Storage System for researchers which is also in the process of being upgraded to include:

- Large SAM-QFS cluster managed HSM system for data management
- 2 Sun Opteron V40z providing metadata servers to the cluster file system and management of the HSM
- Large, expandable online disk storage for projects requiring more on-line data
- Migration to tape storage for large or deep storage using existing data robots and updated tape capacity
- A data intensive computational cluster with mounted SAM-QFS file system
- A cluster of systems dedicated to database access for SQL services and fast index searching
- Scalable front end to enable portals for data management
- Scalable high speed data transfer for grid and streaming services

APAC supports access to the data store for Data Intensive projects on a merit basis, with the following current aggregate allocations:

Projects	Requests (TBytes)	Maximum Usage (TBytes)
APAC DI Projects	72.52	39.68
APAC + Partner Projects	78.93	39.93
Total	151.45	79.61

## The National Grid

The main objective of the APAC Grid is to develop, install and operate a grid infrastructure across the APAC and partner facilities to:

- allow users easier access to the facilities through secure mechanisms,
- allow users to access the most appropriate computing and data storage system in the National Facility,
- provide users with grid services to support dynamic virtual organisations, and
- support users participating in national and international research programs.

The National Grid has been designed as an integrated virtual system with interfaces to external systems and grids; such as institutional computing systems, on-line instruments, sensors and data sources. It is also designed for inter-operation with other grids at the institutional, national and international levels.

Researchers are able to see the resources in the APAC National Grid as an extension of their personal work environments; accessible through a common security system and via portal interfaces. In this sense, the National Grid is intended to become an infrastructure that is transparent to most of its users.

The architecture of the National Grid involves a 'gateway' system at each partner site configured to support a range of grid services and to receive and process grid service requests. The aim of the architecture is to:

- limit the number of systems that need grid components installed and managed thereby reducing overall grid management overheads

- enhance security as many grid protocols and associated ports only need to be open between the gateways and as only the local gateway needs to interact with site systems
- support roll-out and control of production grid configuration through the implementation of standardised grid support across all grid sites
- support production and development grids and local experimentation without significant hardware investment through a Virtual Machine implementation where different services and different quality of services are provided on separate grid installations

The APAC Grid program has engineered, installed and operates grid infrastructure at all APAC partner sites. This infrastructure supports:

- internationally accredited PKI certificates, shared myproxy services, user controlled DN to local account mapping and VOMS/VOMRS based project grouping
- remote data staging and job submission to all significant compute resources within the partnership using gram2, gram4 and LCG protocols
- direct and third party transfers between all sites using GridFTP
- multi-site shared data using federated SRB servers
- portals based on Gridsphere and related tools

The APAC grid middleware stack is based on the Virtual Data Toolkit used in the US Open Science Grid (which incorporates the Globus Toolkit). This deployment has provided the basis for conforming to the emerging international standards for grid services and for interoperation with large-scale grids being developed for researchers in other countries.

It has also facilitated participation in international forums in order to contribute to the development of grid middleware and to gain experience with the early implementation of standards. The participation is fostering cooperation on grid technologies, standards and applications with groups in the US, UK and Asia-Pacific.

Examples of application services developed include:

- workflow scheduling for genome annotation at Murdoch University (not an APAC site) computing using resources at iVEC and SAPAC
- workflow scheduling for geomechanics from iVEC to resources at iVEC and CSIRO HPSC, staging data using federated SRB
- LCG-managed job submission from Melbourne University (not an APAC site) to resources at VPAC, SAPAC, ac3 and ANU
- Belle experiment management using federated SRB to import data internationally, the National Grid to manage jobs and SRB to export data internationally
- digital repositories and portal access solutions for marine and astronomical data sets
- Web-based access to chemistry packages for molecular modelling.

The following application and infrastructure projects are in progress using the APAC Grid.

#### Astronomy

- Astronomy Data Warehouse: Katherine Manson, University of Melbourne. Providing efficient, standardised access to key Australian and international astronomical data collections, providing the ability to discover and query archives and to extract data from those archives.
- Astrophysics Portals: Sarah Maddison, Swinburne University. Providing user-level services for job configuration, submission and monitoring using Gridsphere portlets for theoretical astrophysics codes.
- Gravitational Wave Research: Susan Scott, ANU. Providing the Australian Consortium for Interferometric Gravitational-wave Analysis (ACIGA) to establish a National Gravitational Wave Research Grid which to enhance the collection and analysis of data within Australia, the sharing of this data with international collaborators, and the enhancement of data access and retrieval from these collaborators.

### Bioinformatics

- Bioinformatics Grid Services, Kimberly Begley, Griffith University. Deploying grid-based services for local implementations of major overseas databases and their associated tools.
- Genome Annotation: Matthew Bellgard, Murdoch University. Providing a grid-enabled Blast system; installing a local Ensembl database and the relevant compute/data intensive tools and developing a distributed Genome Annotation System using the rice genome as a model.
- Molecular Docking: Rajkumar Buyya, University of Melbourne. Providing Web-based services to molecular docking applications and a wide variety of chemical databases, as well as tools to assist in the analysis and archival of the screening results.
- Parasitology Grid: Tim Langtry, University of Technology Sydney. Providing services for the sharing and remote processing of biological data for the ARC Network for Parasitology, by adapting tools developed in the European myGrid program (in particular Taverna).
- pBlast: Albert Zomaya, University of Sydney. Providing grid-based services for Web-based BLAST searches, query and annotation commands and bioinformatics portals.

### Chemistry

- Chemistry Portal: Andrey Bliznyuk, ANU. Providing access to advanced computing systems for the computational molecular science community through portal development.

### Earth Systems

- Digital Libraries and Grid Computing: Glenn Hyland, University of Tasmania. Enabling Australian data sets related to oceans, atmospheres, Antarctica and the climate to be made transparently available to the Australian and international research community using grid protocols and standards.

### Geosciences

- Earthbyte: Dietmar Muller, University of Sydney. Providing the foundation for grid-based data access and earth process modelling by linking geological and geophysical observations to palaeogeographic models for constraining mantle convection and lithospheric deformation.
- Geosciences Workflow: Robert Woodcock, CSIRO. Providing workflow services for the geoscience community to access geoscience data sets and modeling software.

### High-Energy Physics

- Belle and ATLAS Experiments: Glenn Moloney, University of Melbourne. Deploying a common data grid infrastructure allowing participation in two international frontier experiments in High Energy Physics: the Belle experiment at the KEK laboratory in Japan; and the ATLAS experiment at the CERN laboratory.
- International Lattice Data Grid: Paul Coddington, University of Adelaide. Providing services for the International Lattice Data Grid (ILDG), that aims to allow highenergy physicists to publish, locate and access hundreds of Terabytes of data from computationally intensive lattice Quantum ChromoDynamics (QCD) simulations.

### Infrastructure Projects

- Computing Infrastructure: David Bannon, VPAC. Deploying the computing systems infrastructure required for an operational APAC Grid, including services for authentication, authorization and job management.
- Information Infrastructure: Ben Evans, ANU. Providing services for an integrated national data grid including data transfer, data federations, information services, and access control.
- User Interface and Visualisation Infrastructure: Raj Chhabra, Queensland University of Technology. Enabling easy access to the National Grid through application portals and visualisation tools for multiple simulations and large-scale data sets.
- Collaborative Support Infrastructure: Chris Willing, University of Queensland. Enhancing Access Grid functionality by adding visualisation and usability enhancements, in order to support collaborative working.

## Appendix VI State based APAC Partners

The following pages contain information relating to the state based APAC partners, which have grown over the last three years into a substantial network capable of regional e-Research infrastructure support.

These are:

New South Wales	ac3
Queensland	QCIF
South Australia	SAPAC
Tasmania	TPAC
Victoria	VPAC
Western Australia	iVEC

## ac3's Capabilities to support NCRIS

### **Background**

**ac3** is an advanced computing centre, established in 2000 by the NSW State Government, and 8 NSW-based Universities. **ac3** operates as an independent company jointly, with the NSW State Government holding 57 per cent equity, and with University partners holding the remainder.

**ac3** is required to operate as a self-sustaining business. The bulk of its business is derived from the provision of managed services and co-location for government and the private sector.

### **Generic Services provided by ac3**

#### **Co-location services**

**ac3** operates two highly secure, highly available, and well connected data centres:

- **Security:** the ATP data centre is accredited to the 7799 data centre security standard, and as such **ac3** is entrusted by quite a few organisations to host sensitive databases.
- **Availability:** **ac3**'s power and air-conditioning infrastructure has a high degree of redundancy, with virtually all single points of failure eliminated.
- **Communications:** dual redundant gigabit links to two different Optus exchanges, which are managed by Optus on a 24x7x365 basis; connections to AARNet, CeNTIE.

#### **Managed Services**

**ac3** provides managed services on a 24x7x365 basis to over 70 organisations, including six of the state's universities — Sydney, UNSW, UTS, Macquarie, Wollongong and Newcastle. These services are based around the ITIL framework, and include:

- Incident Management
- Change Management
- Backups
- Supply of hardware
- Operating system support (Solaris, Windows, Linux)
- Asset Management
- Database administration
- Network design and management
- Security consulting and management
- Storage design and management
- Business Continuity
- Project Management

**ac3** is well placed to serve the research community where high levels of security and availability are required.

### **Support for eResearch**

**ac3** services a broad range of research communities. While the bulk of the support that is presently provided is computationally oriented (involving the parallel computing systems at **ac3** and the APAC NF), our participation in the APAC grid program has fostered a growing range of activity in the data arena. The disciplines supported are:

- **Computational services:** chemistry, physics, materials science, nanotechnology, photonics, engineering, earth systems science, biosciences, geosciences, mathematics, high energy physics,
- **Data services:** bioinformatics, astronomy, geosciences, earth systems science.

In broad terms, the services provided to users comprise:

- Systems administration: managed services, collocation, software installation and maintenance,
- User support and helpdesk facilities for users of local (**ac3**) and national (APAC) HPC systems,
- Computational services: parallelisation, programming development and support, scripting,

- Data services: data storage, management, curation, distributed access (SRB), portal development, workflows, OpenDAP services, GIS services,
- Visualisation services: VTK, ARCGIS, etc
- Access Grid: operation support for users (at partner sites)

**ac3** has key relationships with a number of organisations including:

- 30 NSW Government agencies — contractual relationship for managed services.
- ARC Centres of Excellence—Centre for Ultrahigh-bandwidth Devices for Optical Systems and Centre for Free Radical Chemistry and Biotechnology
- ARC Networks—Parasitology, Earth Systems Science, Molecular and Materials Structure
- SIRCA — Securities Industry Research Centre of the Asia Pacific
- CMCRC — a member of the Capital Markets Cooperative Research Centre
- ANSTO—Australian Nuclear Science and Technology Organisation (just commencing).

## ***Research Facilities operated by ac3***

### ***January 2004***

#### **Computational Facilities**

- Dell Cluster: 152 node (308 CPU) Intel Xeon based Beowulf Cluster, 308GB RAM (total), 1.09TFlops.
- SGI Origin 2400, 64 CPU, 32GB RAM (50GFlops peak).
- SGI Power Challenge, 28 CPU, 4GB RAM (11GFlops peak).
- NEC SX5, 2 CPU, 12GB RAM, 16GFlops.

#### **Storage**

- 1.2 TB RAID 5 Storage.
- IBM 3494 tape silo, 4\*Magstar 3590 tape drives (4TB total).

#### **Network**

- 1 Gbit connection to GrangeNet.
- 1Gbit to CeNTIE
- 1 Gbit link to Optus

### ***Early 2007***

#### **Computational Facilities**

- Dell Cluster: 186 node (376 CPU) Intel Xeon based Beowulf Cluster, 376 GB RAM (total), 1.2TFlops.
- SGI Altix 4700, 16 CPUs, 1.6GHz, 32GB RAM.
- NEC SX5, 2 CPU, 12GB RAM, 16GFlops.
- New capability system to be installed in early 2007: 256 processor, 512GB-1TB RAM, approx 2 TFlops (allocated funds \$1.1M, RFP in preparation)

#### **Storage**

- 3.2 TB RAID 5 Storage.
- 1\*60 Tape silo, 3\*LTO2 drives, running through TSM (6TB total).

#### **Network**

- 1 Gbit connection to AARNet3.
- 1 Gbit to CeNTIE
- Redundant 1 Gbit connections to Optus

# QCIF contributions to NCRIS Platforms for Collaboration

(B. Pailthorpe, Oct 30 2006)

QCIF Ltd is an independent company that provides cyber-infrastructure and related services in Queensland. Its founding members are six of the Queensland universities (UQ, QUT, Griffith, JCU, CQU and USQ). In May 2006, the organisation changed its name from the long-used QPSF.

The mission of QCIF is to deploy and support advanced computing infrastructure to support Queensland's Researchers and Industry.

QCIF directly supports 335 researchers across some 80 R&D projects, with another 100 researchers using related HPC resources (p2). It provides data services to QFAB and the ACcESS MNRF. Some projects are documented at: <http://www.qcif.edu.au/research/>.

## Applications domains

User support spans Computational Science and Computational Engineering, with a focus on:

- Bio-sciences, bio-informatics, computational chemistry
- Engineering & geo-sciences,
  - with particular application to mining industry and associated supporting engineering enterprises
  - computational earth systems science is with links to ACcESS MNRF and pmd\*CRC.
- Ecological sciences (including marine and tropical sciences), with emerging strength in eco-informatics,
- environmental sensing (SensorNets), in collaboration with AIMS.

QCIF operates a Technology Diffusion program that has focused on sustained engagement with a broad range of Queensland industries, including SMEs, and joint initiatives with State Government Departments. The program has operated by seeding relatively small demonstration projects (< \$50k; duration 3-6 mo; start-up in 2-4 wks) that are of limited scope, agile, and able to respond to near-term industry opportunities. Some projects are highlighted at <http://www.qcif.edu.au/industry/>. The program will double in scope from 2007 and will adopt a more systematic approach to engaging with a broader range of SMEs.

## e-Research

QCIF researchers host 11 of the current 37 ARC e-Research grants, and are engaged with a number of related ARIIC projects (eg. DART, ARCHER, MAMS, APSR). QCIF staff initiated the Access Grid in Australia (as well as on the US west coast) and leads the national support and R&D (jointly with ANL; and with CalIT2 for HDV) efforts. QCIF supports data services, including SRB, and leads the APAC Portals development effort. QCIF is the nominated cooperative vehicle for e-Research support in Queensland.

## Scientific data

QCIF is one of the initial pillars of an emerging Australian scientific Data Grid - at UQ, ANUSF, Monash, JCU, UTas and UWA, (along with CSIRO/BoM) – in that each already have in place quite substantial data archival capacity (eg. StorageTek tape robots, staff & emerging data services, etc), funded by universities, ARC, APAC, QCIF, etc. This combined data capacity is approximately 5,000 TeraBytes (TB), with > 500 TB under active management.

We, being the first three sites at least, are now working in a coordinated manner, having won two successive ARC LIEF grants to provide upgrades of ~ 100 TB each pa. Those 3 sites are installing



high performance data access servers in 2007. QCIF directly supports data needs of IMOS (NCRIS 5.12 & AIMS), QFAB and ACcESS MNRF.

### Services:

- HPS systems support; grid AAA; data archives & support, SRB, Workflows, Portals, User Interfaces; SensorNets & Tele-instrumentation (CIMA)
- APAC NF user support, Parallel programming (MPI), Specialist user support for bio-informatics, geo-sciences, chemistry, physics
- Access Grid support & leads Australian AG development effort

### HPC Resources:

QCIF and partners' HPC and supercomputers are distributed across 4 sites, comprising a total of ~1,300 processors with an aggregate peak speed of ~ 4 Tflops.

UQ: sgi *Altix* 3700 supercomputers (132-proc); StorageTek PowderHorn Tape archive (populated at 220TB, with 1,200TB capacity); 15 TB disc; sgi DMF HSM; ViSAC visionarium, 5 x Access Grid nodes; along with the ACcESS MNRF *Altix* (208-proc); three clusters (128- 256- & 100-proc – for Chemistry, Physics, IMB) and a IBM P-690 data engine at IMB. APAC grid gateway.

Griffith 2 SUN clusters (128-proc opteron, 64-proc UltraSPARC), 60 TB disk; 3D wedge VR system; 2x AG nodes (Nathan & Gold Coast).

QUT: sgi Origin 3400 (128-proc), 5 x AG nodes (3 in Gardens Point; Kelvin Grove, Caboolture campus)

JCU sgi *altix* (94 proc), SUN clusters (156 proc Opteron), 4 x AG nodes (3 in Townsville, Cairns); 95 TB StorageTek L180 Silo with DMF.

Other regional AG nodes are located in Rockhampton, Gladstone and Toowoomba (CQU & USQ).

Currently QCIF is upgrading its data capacity at 100TB pa (tapes, matched ~10% disc), often with substantial ARC LIEF funding. The QUT system is undergoing a >100% upgrade in 2006, while the UQ will likely have a 200% upgrade in 2007; aggregate compute capacity is expected to double in the next 2 years.

Major data holdings are for: bio-informatics, spectroscopy (MRI (CMR) and presently crystallography), geo-sciences, marine sciences, satellite imaging, health sciences (MRI, immunology), with data services to AusVO and social sciences.

### Strategic Relationships include:

Qld Govt: Development of State Development, Dept Natural Resources, Mining and Water; Information Industries Bureau, QFAB.

University Institutes: Institute of Molecular Biosciences (IMB - UQ), Centre for Molecular Simulations (CMS - UQ), Centre for magnetic resonance (CMR, UQ), Australian Institute for Bioengineering and Nanotechnology (AIBN – UQ), Centre for Glycomics (GU), Institute for Health and Biomedical Innovation (QUT), ARC Centre for Excellence in Coral Reef Studies (JCU), ARC Centre for Bioinformatics (UQ), ARC Centre for Complex Systems (UQ), pmd\*CRC , Railways CRC (CQU & QUT), Creative Industries CRC (QUT).

National Organisations: ACcESS MNRF; Australian Institute of Marine Science (AIMS); AusVO, MTSRF (successor to the Rainforest & Reef CRCs); Molecular and Molecular Sciences Network (ARC Research Network); Sensor Networks (ARC RN); DART (ARIIC Project); Australian Institute for Commercialisation (AIC).

International: San Diego Supercomputer Centre (SDSC) – SRB group (R. Moore); Argonne National Lab – AG & Display groups, MCS Division; Indiana U (R McMullen – CIMA); CalIT2 / SIO – SensorNet projects, Data/Control rooms (HDV); NSF – participation in US planning meetings (both via AG and on-site); NSF-NARA working party for the preservation of scientific data; PRAGMA - co-hosted (with APAC) 10<sup>th</sup> meeting; UK e-Science program.



# South Australian Partnership for Advanced Computing

**The role of the South Australian Partnership for Advanced Computing (SAPAC) is to provide e-Research support to the South Australian research community through the provision of the full spectrum of e-Research infrastructure, services, training and expertise.**

SAPAC is an unincorporated joint venture of the three South Australian universities – Flinders University, the University of Adelaide, and the University of South Australia. It first began operations in October 2002 and was formally constituted as an unincorporated collaborative venture on December 12, 2003. SAPAC has built on the prior work of the S.A. Centre for Parallel Computing (a three university collaborative venture that continuously operated from 1991) and the Centre for High-Performance Computing and Applications (at the University of Adelaide).

## **Major research groups supported:**

- Australian Centre for Plant Functional Genomics
- Cooperative Research Centre for Molecular Plant Breeding
- Adelaide Proteomics Centre
- Special Research Centre for the Molecular Genetics of Development
- Special Research Centre for the Subatomic Structure of Matter
- Ian Wark Research Institute, UniSA
- Applied Centre for Structural and Synchrotron Studies, UniSA
- Institute for Telecommunications Research, UniSA
- Teletraffic Research Centre, University of Adelaide
- Computational chemistry groups at University of Adelaide and Flinders University
- Physical oceanographers at Flinders University
- South Australian Museum
- Australian state, territory and national herbaria

Major research areas supported include life sciences, high-energy physics, synchrotron studies, chemistry, materials science, physical oceanography, geology and geophysics, parallel and grid computing research, turbulence and combustion modeling (computational fluid dynamics), telecommunications network analysis, environmental science and biodiversity informatics.

## **e-Research Services offered:**

- High-performance computing; Grid computing infrastructure and services; Grid applications development;
- Information infrastructure and data management (e.g. the South Australian Sustainable Repository);
- Visualisation, haptics and virtual reality;
- Access Grid and remote collaboration toolsets; Remote access to instruments;
- Portals and application-specific tools for research communities;
- Induction and specialised training;
- Helpdesk and user support services;
- Specialist e-Research support (e.g. bioinformatics);
- Contract R&D and consulting to industry and government;
- Outreach, e-Research and research promotion; Seminars and workshops.

Some examples of specific support to research areas and research groups are:

- **Bioinformatics:** As part of the APAC grid bioinformatics project, development of software and web portals for standard bioinformatics jobs such as BLAST, ClustalW and HMMER. For wider applications we also provide customised portals, workflows as well as grid and database support. Bioinformatics is one of the main growth areas at SAPAC.
- **High-energy physics:** development, optimization, documentation and improved software engineering of parallel applications for the Centre for the Subatomic Structure of Matter, which is one of the largest users of SAPAC facilities and the APAC National Facility.
- **Materials Science:** ARC e-Research project to provide remote monitoring, and real-time data download, storage and processing of data, for synchrotron experiments; working with national and international groups. The expertise gained will be applied to other instruments.
- **Biodiversity informatics:** a new version of Australia's Virtual Herbarium, a federated database system linking plant specimen databases of all the major Australian herbaria. SAPAC will also participate in the Atlas of Living Australia project.
- **Research Data Repositories:** data repositories for high-energy physics, synchrotron experiments, bioinformatics and biodiversity informatics, and humanities and social sciences; including the South Australian Sustainable Repository (SASR) project to support SA research data management needs.

#### **Some data on SAPAC users and facilities:**

**Users:** SAPAC currently has around 375 registered users (120 active users of supercomputer facilities in 2006H1), up from 100 registered users in January 2004.

**Staff:** SAPAC currently has 17 paid technical and administrative staff (11.5 FTE), compared to 5 paid staff (3.5 FTE) in January 2004.

Several academic staff and students also work on SAPAC projects, providing an additional in-kind contribution. SAPAC also pays several summer interns to work on various projects.

#### **Supercomputer development:**

<i>January 2004</i>	A total of 1.45 Tflops, 548 CPUs, 5 Tbytes RAID storage
<i>October 2006</i>	Aquila added to give a total of 2.28 Tflops, 708 CPUs, 20 TBytes RAID storage
<i>1<sup>st</sup> Quarter 2007</i>	With new ARC LIEF-funded SA Supercluster Facility will provide a total of ~6 Tflops, ~900 CPUs, ~30 TBytes RAID, plus robotic tape silo with > 40 TBytes

#### **Description of individual supercomputer facilities:**

- *SA Supercluster Facility* – Architecture to be decided but approximately 3.5 Tflops, around 340 to 400 CPUs, Infiniband interconnect. To be installed in early 2007.
- *Aquila* – SGI Altix 3000, 0.83 Tflops, 160 Itanium 1.3 GHz CPUs. Installed August 2004.
- *Hydra* – IBM eServer 1350 Linux Cluster, 1.2 Tflops, 256 Xeon 2.4 GHz CPUs, Myrinet interconnect. Installed June 2003 and was second fastest computer in Australia and #106 in Top 500 list.
- *Orion* – Sun Technical Compute Farm, 0.14 Tflops, 160 UltraSPARC II 450MHz processors, Myrinet network. Installed June 2000 and was fastest computer in Australia and #188 in Top 500 list.

#### **Other facilities and infrastructure for e-Research support:**

SAPAC manages two stereo 3D visualization facilities, the South Australian Virtual Reality Centre (SAVRC) and the SAPAC VisLab, as well as a haptics lab. The SAVRC, SAPAC VisLab and the SAPAC meeting room all have Access Grid facilities; all developed since January 2004.

SAPAC is developing the South Australian Sustainable Repository, a federated distributed research data repository for the three SA universities.

## University of Tasmania’s contributions to NCRIS Platforms for Collaboration

### Vision – Collaborative Platforms for the Marine and Earth Systems Science Research community.

- Enabling Australian data sets related to oceans, atmospheres, Antarctica (and climate) (TPAC) and marine data sets (BlueNET) to be made transparently available to the Australian and world research community using emerging international standards (including grid technologies).
- Providing the collaborative infrastructure to enable researchers to more rapidly observe, analyse and simulate changes in the ocean environment, including high speed networks and high performance computers

The systems we are implementing are key enabling technologies for e-research in the marine arena, facilitating ready discovery of, access to, and visualisation of marine data.

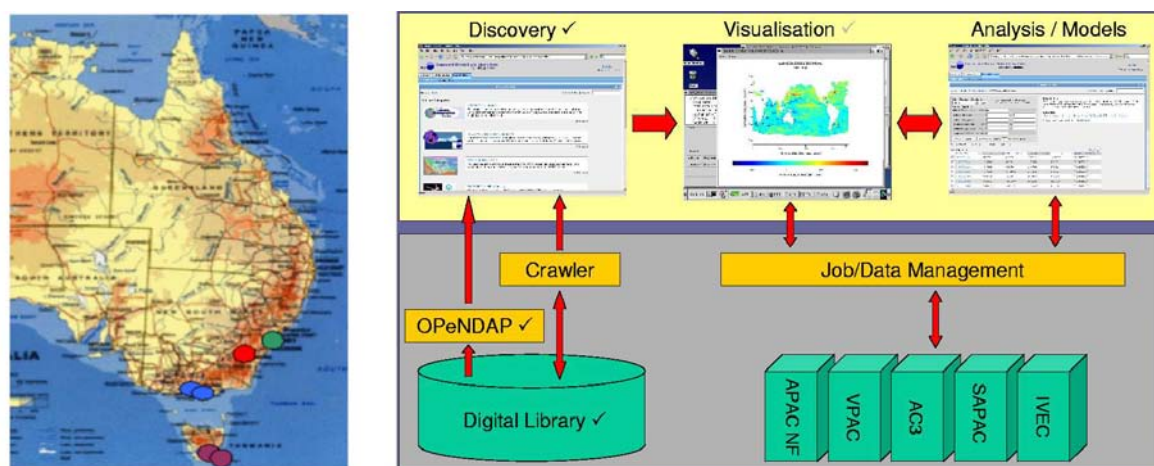


Figure1. Shows the TPAC distributed digital repository of climate and ocean simulations, and an illustration of the knowledge discovery portal, visualization and analysis tools.

Global Collaborations	National Collaborations
<p>The TPAC digital library is a member of the National Virtual Oceanographic Data Server (NVO DS).</p> <p>UK e-Science program (including collaborations with the BODC and BADC)</p> <p>OpenDAP Inc (developers of OpenDAP)</p> <p>Geophysical Fluid Dynamics Laboratory, NOAA, Princeton University</p>	<p>The ARC Research Network proposal for Earth Systems Science, a community of 200 members.</p> <p>AODC-JF (Australian Oceanographic Data Centre –Joint Facility)</p> <p>MAMS (Meta Access Management System)</p> <p>ASK_OSS</p> <p>OAK Law (Open Access Knowledge Law)</p> <p>APSR (Australian Partnership for Sustainable Repositories)</p> <p>APAC (Australian Partnership for Advanced Computing)</p> <p>ACCESS (Australian Climate Earth System Simulator)</p> <p>AusCOM (Australian Community Ocean Model)</p>

### Partners

TPAC Partners	BlueNET partners
<p>ACE CRC (Antarctic Cooperative Research Centre)</p> <p>CSIRO Marine Atmospheric Research</p> <p>BMRC (Bureau of Meteorology Research Centre)</p>	<p>University of Queensland,</p> <p>James Cook University,</p> <p>University of Sydney,</p> <p>University of Melbourne,</p> <p>University of Adelaide,</p>

AAD (Australian Antarctic Division) Australian Maritime College APAC (Australian Partnership for Advanced Computing) ARC Network Earth Systems Science	Flinders University, University of WA, University of Tasmania [lead institute] Australian National University Bureau of Meteorology GeoScience Australia CSIRO Marine and Atmospheric Research Australian Antarctic Division Royal Australian Navy Australian Institute of Marine Science
---	--

### **University of Tasmania Facilities**

The University of Tasmania is making significant investments in high-performance computing, data storage, and high-speed networks within Tasmania (via the Tasmanian Research Education Research Network). Current Facilities are (June 2006 and January 2007)

- 128 processor SGI 4700, 320 Gbytes of RAM (0.8 Teraflops)
- 50 Terabytes of Raid 5 Disk Storage
- 1 Tape Silo, 66 Terabytes (mirrored) under HSM through DMF and 4 processor SGI 350
- Tasmanian Research Education Network (TREN) (Currently 155 mbps, ~1Gbps from University of Tasmania to AARNET across Bass Strait in late 2007).
- Total users (~50 active staff and students at University of Tasmania, Australian Antarctic Division, and CSIRO with direct access to these facilities. Over 150 user accounts)

The University of Tasmania (TPAC) has renewed the HPC facilities 4 times over the 14 years, and had one incremental upgrade. Each major upgrade has had a budget of ~\$1M. The Robotic silo system has been upgraded and grown with a separate budget.

The Facility is open to users nationally as well as within the TPAC partnership. This national use has been much improved by the use of APAC certificates and connection of our facilities to the APAC national grid. There have been consultancies also carried out for Tasmanian Government Businesses and Industry.

TPAC Services to users:

- Generic systems administration and user support and helpdesk facilities
- Access Grid node for University and TPAC partner use
- Support for users of APAC national facility and the APAC National Grid
- Support and expertise for Earth Systems Science models
- Support for expertise for Earth Systems Science data repositories
- development of portals to support above activities and the use of the National grid.

### **Links:**

TPAC Digital Portal for Earth System Science: <http://ngdev.sf.utas.edu.au:8080/gridsphere/>

National Virtual Oceanographic Data Server: <http://www.nvods.org>

### **Contacts:**

Assoc. Prof. Nathan Bindoff, Director TPAC, [n.bindoff@utas.edu.au](mailto:n.bindoff@utas.edu.au), (03) 62262986, (03) 62267888 (messages)

Prof. Craig Johnson, Chief Investigator, BlueNET, [craig.Johnson@utas.edu.au](mailto:craig.Johnson@utas.edu.au), (03) 62262582

Kate Roberts, Director BlueNET, [Kate.Roberts@utas.edu.au](mailto:Kate.Roberts@utas.edu.au), (03) 62262873

### Background

VPAC is an independent company, founded in 2000 by 6 Victorian Member Universities, with \$6M of seed funding from the Victorian State Government. VPAC is a founding Partner of APAC. VPAC's business model is that the Member Universities contribute fixed annual subscriptions in proportion to their research needs for Advanced Computing services and support. In return, VPAC delivers research services and support to Members under a resource distribution model that targets ensuring that all Members receive an equitable return on their investment. VPAC supplements its income from Member subscriptions from a variety of funding sources, notably from provision of contract R&D support to industry and other organizations. Profit from such supplemental income is returned to Members through a competitive grant scheme.

VPAC has grown to point where, in November 2006:

- All 8 Victorian Universities are Members of VPAC, through 2011, with total subscription of ~\$2.0M p.a.
- VPAC's contract R&D income is approximately ~\$5.0M p.a., generating a profit of ~0.5M p.a., with approximately 30 current R&D contracts for 20 clients
- VPAC employs ~50 staff in two locations
- VPAC has well over 400 registered users of its HPC facility, of which ~200 are active

### Vision

VPAC's vision is to be a leading international e-Research service provider, serving the needs of its Member Universities, the State, and national and international communities.

### Strategy

VPAC's strategy is based growing its collaboration, resources, and business in e-Research support. The key themes of that Strategy include:

- Maintaining strong relations and collaboration with all VPAC's stakeholders, including Member Universities, the State government, and APAC
- Support the evolving R&D support needs of Members of VPAC, through provision of a range of services and facilities - notably HPC, data storage, data and compute grids, and visualization
- Growing VPAC's national and international collaborations and R&D support reputation
- Continuing to grow VPAC's commercial R&D support activities, and broadening VPAC's range of clients – while continuing to focus on Advanced Computing R&D areas that complement Member and State strengths

### Facilities – Computational, Data, and Visualization

VPAC has a long-term strategy of operating two significant compute clusters and replacing one of them every two years. This is a cost effective means of purchasing compute capacity and means:

- Each machine is replaced every four years and VPAC always has one machine less than two years old.
- The end users are provided with an easy and relaxed migration path.
- There is significant capacity available, even when decommissioning an existing machine and installing a new one.

VPAC currently operates two large High Performance Compute systems and several smaller systems for specific purposes. Total peak capacity is around two GFlops (~450 CPUs) and there are plans to replace the x86 system early in 2007. VPAC's remaining systems are all 64bit, being IBM Power5 and Opteron based. VPAC achieves a very high usage rate, usually above 80% for its HPC systems; this figure is not adjusted in any way for scheduled or otherwise downtime or system unavailability. For its two main machines, the usage is typically in the high eighties. There is thus a high and growing demand for HPC – since 2000 VPAC's compute capacity, measured in FLOPs, has grown by a factor of x10.

Additionally, VPAC operates a number of smaller, but not insignificant, clusters for other organizations, particularly two for Member universities and two for groups with whom commercial arrangements exist. VPAC operates a consolidated user file system so that replacement of a compute system does not require users (or VPAC systems

administrators) to move user files from an old system to a new. VPAC does sell time on its computer clusters for commercial use, but this is a relatively small amount. Most commercial users are either interested in full-service solutions (from modeling to visualization) or in advice and consultancies on scoping or purchasing HPC facilities for their own use.

The total storage that VPAC manages is ~12Terabytes, with some large datasets being hosted for Members. VPAC operates a Visualization facility for a commercial client.

#### Demonstrated Technical Strengths

VPAC services and projects are operated on a fully commercial basis, by full-time staff, and VPAC's business processes have been ISO 9001 accredited for over 3 years. Increasing, e-Research and Advanced Computing require a very diverse and collaborative skill base, ranging from software engineering and systems administration to scientists and engineers.

VPAC's employment profile consists roughly of:

- 10 Systems administration and systems programmers
- 10 Computational engineers
- 8 Software engineers
- 2 Scientists – biochemistry and geospatial sciences
- 5 Project managers

Aside from this pool of employees, VPAC has close technical ties with and collaboratively works with many more technical staff at Member Universities. The annual growth in VPAC staff is about 20% p.a. with a very low turnover. VPAC runs a highly competitive summer internship program, which has employed an average of about a dozen students p.a. each year for the past 6 years (in 2006 there were over 200 applicants). VPAC also runs a strong technical training program, both for Member staff and for other research organizations.

VPAC has projects, expertise, and staff actively working in the following technical areas:

- Engineering and Manufacturing (CFD, FEA, CAD, and Virtual Engineering)
- Life Sciences (Computational biochemistry and medical informatics)
- Geospatial Sciences (Remote sensing and geospatial modeling and visualization)
- Computational Science (Scalable parallel solvers, solver frameworks and geosciences applications)
- Grid Computing (Authentication and authorization; middleware; portal deployment)

VPAC manages the Compute Infrastructure of the APACGrid and has delivered a range of success stories that are internationally recognized (e.g., use of virtualization for grid gateways).

#### Major Clients and Projects

VPAC has ongoing projects and collaboration, funded from sources that include

- California Institute of Technology (Caltech) [School of Earth Sciences]
- Holden and GH (North America)
- Caterpillar
- 2 major international aerospace companies
- BHP Billiton
- Melbourne Health
- Victorian Infectious Diseases Reference Laboratory (VIDRL)
- VICNISS – Hospital Aquired Infection Survaillance
- Falls Creek Alpine Resort
- Department of Primary Industry; Department of Sustainability and the Environment; Country Fire Authority
- Bureau of Emergency Services
- Life Saving Victoria

#### Strategic Relationships

VPAC has strategic relationships and agreements with many organizations, but the following are noteworthy:

- APAC and its Partners
- The Victorian State Government and at least 4 of its departments
- E-Research groups in the USA (VPAC has a staff member on the Executive Board of the NSF funded Center for Computational Infrastructure in Geodynamics ([www.geodynamics.org](http://www.geodynamics.org)))
- General Motors
- New Zealand (Canterbury University)
- PRAGMA and OGF Working groups

## Background

iVEC has been providing advanced computing services to the WA research community since its inception in 2000. The iVEC partnership includes the three research intensive universities in WA, CSIRO and Central TAFE and has received over \$5.2 million of state funding to date. In May 2006, the WA Treasurer Eric Ripper announced that the State Government had set aside an additional \$1.95 million a year over the next four years to continue funding iVEC into the next decade. In the business plan for this funding, iVEC has committed to delivering access to its physical infrastructure via grid technologies and e-Research services and is looking to maximise the state government investment by partnering with commonwealth funding agencies in the development of its physical and human infrastructure.

## Strategy Going Forward

iVEC's strategy includes:

- The development of first class e-Research infrastructure in Western Australia via the installation of a petabyte scale data storage facility and a series of significant upgrades to the computational, networking and visualization facilities;
- Increasing the human capital necessary to both ensure seamless integration of iVEC's facilities into the National Grid and provide quality user support to users of the iVEC systems; and
- Integrating State key clients, including Government and Industry, into iVEC programs.

Through these steps, iVEC sees demand for these services by the WA research community growing rapidly over the next five years.

## Current Application Domains

The major user communities of iVEC are Astrophysics (Gravitational Wave and Radio Astronomy), Bioinformatics, Computational Biology and Biochemistry, Computational Material Science, Computational Chemistry, Computational Geoscience, Computational Physics, Genetic Epidemiology, Remote Sensing, Water Research, and eBusiness.

Many of these communities are to be supported by NCRIS Capability areas. Some notable examples include:

- Radioastronomy, where WA is one of the two shortlisted sites for the \$1.6 billion Square Kilometer Array radio-telescope. Three Professors in Radioastronomy have been appointed in WA, who will be involved in creating the advanced technology required to build such a software telescope as well as undertake the large scale simulations required to understand the results from the telescope.
- Bioinformatics, where iVEC will invest \$300k in 06/07 for the formation of an Informatics Facility. W.A. has also spearheaded the APAC Grid Bioinformatics Project.
- Computational geology, where iVEC possesses the expertise to develop the geoscience community specific interoperability standards required along with the ability to implement the necessary grid infrastructure to achieve a functioning geoscience grid.
- Remote sensing, where iVEC is now storing data from live satellite feeds that cover both the land mass of WA and the ocean off the coast. In the past, this data has all been offline but utilising the iVEC infrastructure this data can now all be accessed almost instantly.

## Desired Areas of Strength

iVEC is looking to extend its computational communities to include climate modelling, tsunami modelling, CO<sub>2</sub> sequestration, geothermal modelling, robotics and oil & gas.

## Industry and Government Uptake Program

The Industry and Government Uptake Program is responsible for engaging industry and government agencies to encourage the uptake of grid computing, data storage and scientific visualisation directly and through the adoption of e-Research principles. This program commenced in Jun 2005 and has been very successful to date, undertaking nine joint projects with industry/government worth \$1 million. In addition, over twenty formal and twenty-five informal seminars, receptions, demonstrations and talks have been held to promote iVEC's facilities and capabilities to government agencies and industry.

## Facilities at January 2004

iVEC initially focussed on visualisation and its HPC capacity at this time was restricted to a 16 processor AlphaServer system rated at 25 GFlops. It had 16 GB of RAM and 320 GB of shared disk. The iVEC staff complement was 5.5 EFT and between 2002 and 2004 there were 30 users on 17 projects.

## Facilities at January 2007

iVEC has two primary compute resources totalling almost 2 TFlops:

- A 168 processor SGI Altix system with 336 GB RAM and 10 TB of 500 MB/s disk
- A 164 processor Cray XT3 system with 330 GB RAM and 30 TB of 1.1 GB/s disk

The number of iVEC staff is now 16 and we have 45 project investigators with 127 users.

By June 2007, iVEC will have a third system, a cluster system of at least 1 Tflop bringing the total capacity to 3 Tflops. In addition, a petabyte scale storage facility will be purchased. It is expected that this facility will have a capacity of 0.5 PB at installation and 40 TB of high speed disk at the front end. It is anticipated that this will become a 2 PB system by 2010.

## Strategic Relationships

iVEC has a number of strategic relationships with organizations both within Australia and internationally. Firstly the composition of the iVEC partnership is unique within Australia with both CSIRO and Central TAFE as core partners in addition to the universities. This brings a valuable industrial focus to iVEC as well as outreach and training. In addition, our strong relationship with the Department of Industry and Resources has contributed to our successful refunding.

iVEC has significant strategic relationships with many State Centres of Excellence as well as most of the Cooperative Research Centres headquartered in WA. iVEC is a core participant in three proposed State Centres of Excellence, two in multiscale modelling and one in radioastronomy. Finally, iVEC has signed a MoU with ISA Technologies, an independent consultancy and specialist technical company that provides Supercomputing On Demand as well as other IT related services. The agreement allows iVEC partner organisations to access the facilities at ISA Technologies for specific projects that fall under the auspices of iVEC.

Nationally, iVEC has strategic relationships with the pmd\*CRG, which has computational geoscience as a core program and the AcCESS MNRF. iVEC researchers are driving the SEE Grid, which is bringing together people in the earth, environmental and computing sciences to address the issues of "transparent access" to data and knowledge about the earth, and the available and potential technologies offered by the grid that enhance the ability to explore for and manage Australia's natural and mineral resources. iVEC is also involved in a research consortium with the ATNF. iVEC has built very strong links with the CSIRO HPSC, resulting in them coinvesting in iVEC's HPC resources as well as directly networking the two sites at 1 Gbps.

Internationally, iVEC has formed alliances with the Cray Users Group, the Swiss National Computing Centre, the University of New Mexico, and the Pacific Rim Regional Visualisation and Analytics Center (PARVAC). iVEC is part of the HIT lab New Zealand Consortia and has strong links with the eScience program of the EPSRC in the UK through computational chemistry and geology as well as eBusiness. A group of iVEC researchers are consultants with Accelrys Inc, the worlds leading materials modelling company whilst another is the co-Chair of the International Wheat Genome Sequencing Consortium (USA, France, China). Finally iVEC is collaborating with NASA, porting some of our codes to their Columbia system.