NATIONAL COLLABORATIVE RESEARCH INFRASTRUCTURE STRATEGY

Final Investment Plan for the Research Capability

Platforms for Collaboration

CONTENTS

Main Report

APPENDICES

PART ONE FACILITATOR'S DECLARATION

I, Rhys Francis, confirm that I have prepared this Investment Plan in accordance with the NCRIS principles as set out in the NCRIS Investment Framework.

Signed

Rhys Francis

Name

4 / 4 / 2007

Date

Summary Messages:

The three major additions to the infrastructure landscape appearing in the report are:

- The establishment of the Australian Access Federation (AAF) and the deployment and use of its services across the entire NCRIS investment space (and beyond)
- The establishment of the Australian National Data Service (ANDS), to address the issues raised by the Prime Minister's Science, Engineering and Innovation Council (PMSEIC), and to establish a properly constituted leadership group charged with assisting research institutions to develop and implement strategies for research data management
- The establishment of a tier of infrastructure focussing on capability specific needs rather than generic needs, to bring compute and data investments into alignment over time.

The three major changes to current infrastructure delivery appearing in the report are:

- The re-purposing of the APAC National Grid towards support for researchers across the spectrum and in particular to significantly increase the capability in data and web services technologies
- The extension of the role of regional service providers to include operational support for national inter-operation for major research resources, and particularly resources associated with NCRIS capabilities
- The establishment of an e-Research architecture development process to align research communities, middleware and operational services; and to consolidate the development of discipline focussed 'grids' built by configuring underlying generic services.

Contextual issues that surround these investments and which need to progress with them include:

- The provision of effective access for all researchers to all resources based on the harmonisation of underlying campus, regional and national network and authorisation infrastructures
- The development of policies and regularity environments that enhance e-Research and enable easier research collaboration
- The development of support systems for the growth, enhancement and sharing of e-Research expertise; including rewards, incentives and career recognition; within publicly funded research organisations.

PART TWO OVERVIEW

Modern research is increasingly powered by 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 platforms.

Broadly speaking, this ground work has been recently and expertly covered by the e-Research Coordinating Committee (eRCC)[1], which says:

"Just as in other facets of life, information and communication technologies (ICT) are radically transforming the way research communities across the world are operating. Distributed high-performance computing, digital data resources and high speed communications are just some of the developments improving the capacity of researchers to interact with their colleagues and share data worldwide in ways previously un-heard of.

"These 'e-Research' capabilities are enabling researchers in fields as diverse as medicine, genetics, chemistry, education, linguistics and finance to achieve world class research outcomes and to disseminate knowledge gained from research through the use of advanced ICT. In fact, e-Research has the potential to increase the efficiency and effectiveness of research endeavours across all disciplines. Greater interactivity between researchers and an increased ability to access research outputs will benefit industry, governments and the Australian community as a whole."

The rapid growth in the quantum of research data, and the need for better management of these assets, was recognised by the eRCC as an important issue to be addressed by the research community and governments alike. This issue has also been identified internationally by agencies, such as the US National Science Foundation and the UK's Office of Science and Innovation, and improved e-Research capabilities will be the key to finding a solution.

Consequently, the Platforms for Collaboration Investment Plan has been informed by the considerations of the eRCC; the needs of NCRIS investments; the recommendations of the PMSEIC working group on Data for Science; extensive consultations; and reviews and inputs from existing activities.

[1] An Australian e-Research Strategy and Implementation Framework, Final Report of the e-Research Coordinating Committee, DEST April 2006.

Investment Principles

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

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

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

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

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

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

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

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

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

[VISION]

"Australian researchers will enhance their contribution to world-class research endeavours and outcomes, through the use of advanced and innovative information and communication technologies." [eRCC]

e-Research is the use of advanced and innovative information and communication technologies to support both traditional fields of research, as well as new fields of research made possible by those advanced technologies. e-Research embraces the increasingly team-based nature of successful research. It is vital because it allows research to take a quantum leap forward by enabling new research activities.

e-Research also challenges the notion of what we have traditionally thought of as comprises research infrastructure. As we move from a bricks and mortar environment to a digital environment, our view of infrastructure must change. e-Research infrastructure not only comprises hardware relating to computing, data storage and communications networks, it can be perceived as the complex models developed with considerable investments to examine large research challenges, such as climate change, water management and land use. Such models exist in a virtual environment, but may be regarded as research infrastructure nonetheless and increasingly will need to be recognised by funding agencies responsible for infrastructure investments.

A key reason for government investment in e-Research is that it increases the economic, social and environmental returns to Australia from its research funding. This is because it will:

- increase the return on the Government's substantial investments in both research and research infrastructure;
- ensure Australian researchers remain internationally competitive, including in the emerging fields of research; and
- increase benefits from researcher-industry collaboration.

As e-Research has evolved in the past, the Australian Government has made specific investments in the individual elements that underlie the vision for e-Research in Australia. These include:

- a robust high bandwidth telecommunications network, the Australian Research and Education Network (AREN);
- distributed high performance computing capacity, the Australian Partnership for Advanced Computing (APAC);
- accessible digital repositories (research databases, online libraries, etc);
- accessible research facilities and large instruments (telescopes, etc);
- agreed standards and specifications to maximise interoperability between networks, computer platforms and applications including middleware (the software 'glue' to assure accessibility, etc); and
- the Australian Access Federation (AAF).

However, the research infrastructure landscape in Australia will change in the coming years, principally through the Australian Government's investment in a new generation of research infrastructure. These investments in turn require further investment in e-Research if the vision is to be realistically addressed. These investments will be in the e-Research platforms that provide researchers with the ability to generate, collect, share, analyse, store and retrieve information.

The Platforms for Collaboration capability of the National Collaborative Research Infrastructure Strategy will provide many of the platforms for Australian researchers to engage in e-Research, both domestically and internationally. In doing so, it will support the vision proposed by the eRCC.

The following key investment areas are recommended for the PfC capability:

- National Data Management Infrastructure;
- National Computational Infrastructure; and
- Interoperation and Collaboration Infrastructure.

The investments are predicated on two key services that will provide a foundation for PfC investments. These foundation services are (1) AREN, and (2) the AAF. Given the importance of these foundation services, funding has been provided in each case to establish or equip them to support the PfC capability investments.

While the PfC capability will ensure major steps forward are made, much remains to be done to fully achieve the vision for e-Research in Australia. For example, issues that are beyond the scope of what PfC might address within its current funding envelope include broadband access for researchers to each relevant resource from their desktop/laptop computer; the need for a future upgrade of the AREN backbone, or its international linkages; and, most vitally, sustained funding for the long term storage and management of research data of enduring importance.

Overall Arrangement

On acceptance by the NCRIS Committee of this PfC Investment Plan, the Department of Education, Science and Training (DEST) will establish the Australian e-Research Infrastructure Council (AeRIC) to which direction will be given regarding policy and investments intentions from the NCRIS committee.

AeRIC will develop and sustain the forums in which co-operation can be achieved, thereby, over time, providing policy and standardisation frameworks that deliver a nationally coherent e-Research infrastructure.

AeRIC will recommend the manner in which funds will be provided to PfC components in accordance with the investment plan, and subject to annual re-assessment with DEST.

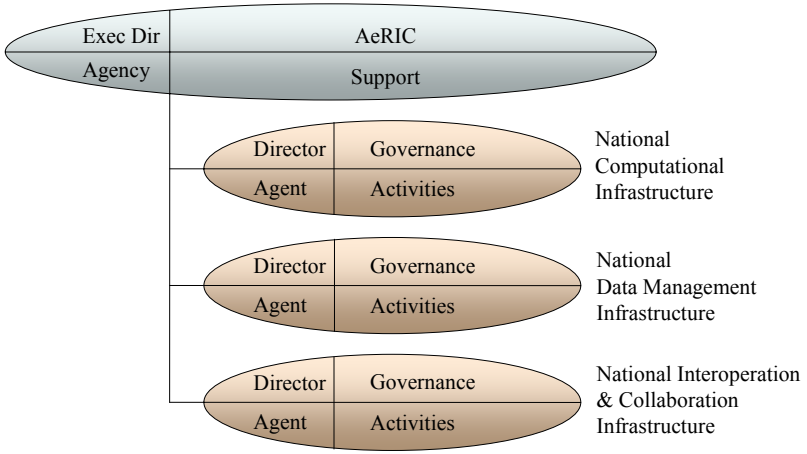
PfC components will report to, and be reviewed for progress by, AeRIC which will recommend variations to funding agreements to the NCRIS Committee when required.

The overall mission of AeRIC will be to ensure that world class services and expertise are identified, developed and delivered nationwide in ways that support effective e-Research within and across all research disciplines. This will include services and expertise related to:

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

Governance

Because of the critical contribution ICT will make to future research, and because the research practice and landscape will continue to evolve in response to increasing ICT capabilities, the activities of AeRIC will have wide ranging ramifications. The members of AeRIC will be determined by DEST in consultation with higher education and research institutions and other relevant agencies.



The Chair will be a person with a broad view of national e-Research requirements.

The Members of AeRIC (about ten members are envisaged) will be drawn from the stakeholders and participants in e-Research infrastructure; including providers and users of AeRIC e-Research services and capabilities.

To ensure the widest possible input, AeRIC will host a broadly inclusive e-Research forum to gather strategic input in refining its business plan and activities.

An Executive Director will be appointed to lead the identification of integrative services of value across NCRIS capabilities and the development of an annual business plan for approval by AeRIC and the NCRIS Committee.

The Executive Director will also be responsible for monitoring implementation arrangements and activities, identifying and managing risks, and measuring the key performance indicators of implementation activities.

A small secretariat will be established to assist the Executive Director. The secretariat will be able to support grant management through DEST.

The PfC components will be the primary means through which a coherent suite of services is developed and delivered to the users of the infrastructure.

National Data Management Infrastructure

The effective re-use of research data on a national basis is the primary goal of the investment into national data infrastructure. The investment will deliver research data registration, location and access services, and outreach services for researchers and institutions that can enhance the effective use of data within a federated research data management system (as recommended to PMSEIC). The investment also meets some key requirements of the NCRIS capabilities:

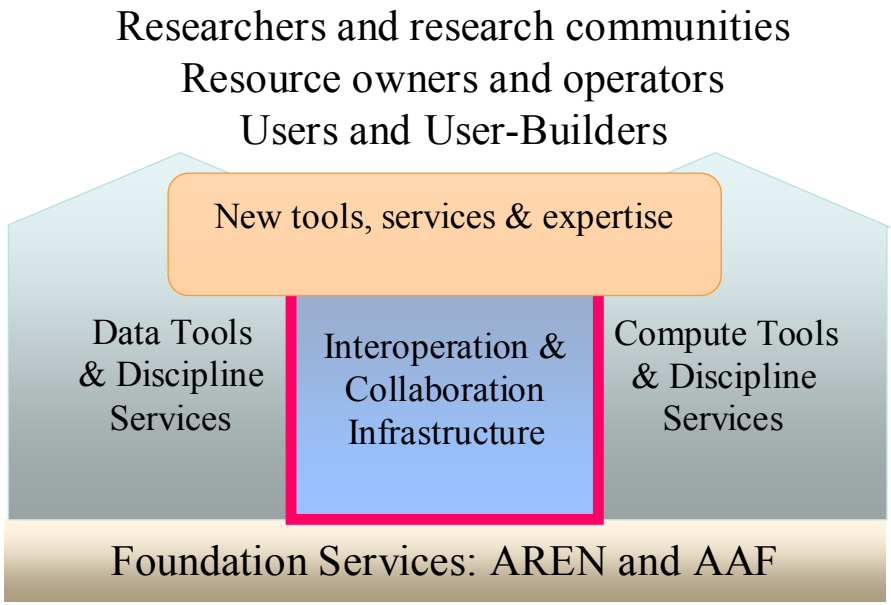
5.1	Evolving Biomolecular Platforms and Informatics
5.2	Integrated Biological Systems
5.3	Characterisation
5.8	Networked Biosecurity Framework
5.10	Optical and Radio Astronomy
5.12	Integrated Marine Observing System
5.13	Structure and Evolution of the Australian Continent

The outcome will be the ability for all researchers to identify, locate, access and analyse any available research data, regardless of origin or scale.

Achieving effective re-use is particularly difficult for collections that are not clearly within the responsibility of any of the organisations which collect and preserve data sets. Therefore the investment will also provide stewardship of some nationally significant data collections, using capability around selected existing research data retention facilities, and aiming at a leadership example of research data management.

In summary, the investment in the national data infrastructure will help users, builder-users and experts improve community data management practice, by providing:

- outreach advisory and support services to assist establish data curation practices;
- outreach advisory and support services to assist institutions establish research data management infrastructure and practice in a nationally coherent manner;
- data management brokering to help match needs to providers;
- federation level services, including data location, access and data mining;
- stewardship services hosting some categories of collections; and
- a merit scheme for prioritising the best use of the limited resources.



National Computational Infrastructure (NCI)

High performance computing (HPC) requirements are explicitly identified in 5.1, 5.10, 5.12, 5.13 and are generally required within the discipline demands related to all NCRIS capabilities. The investments will deliver an internationally significant computing capability which can be assigned on a merit and priority basis, and build essential expertise in HPC needed to support NCRIS-supported and other priority research.

The outcome will be computational modelling beyond the reach of normal project, department and institution investment.

The spread of e-Research is also expected to be accelerated by the propagation of tools and analysis systems (via portals and similar encapsulations). The investment will also develop a computing fabric that supports the migration of tools and systems towards users.

The outcome will be a national computing environment in which computing power can be configured to match the tools and analyses required by priority research.

In summary, the national computing infrastructure will provide those computational services which can best be funded and justified at the national level, by providing:

- an Australian capability system commensurate with international practice;
- advisory and support expertise to assist computationalists;
- oversight and direction setting for the national computing fabric; and
- a community based merit allocation scheme for funded resources.

Interoperation and Collaboration Infrastructure

For e-Research to become easier for new e-Researchers, the complexity of the underpinning ICT systems needs to be hidden; fast, correct and expert help needs to be provided and the co-ordinating systems need to be managed in an expert way.

This investment creates an Australia-wide inter-operation platform linking expertise and compute and data support systems across regional and national service providers into a day-to-day operational activity.

The functions will include:

- a first point of contact and the management of effective user support;
- robust implementation of nominated middleware, tools and portals;
- robust implementation of collaboration services;
- a job submission fabric across all accessible and significant compute resources;
- a data movement fabric between all major data and compute resources; and
- the day-to-day operations management of the above services.

e-Research is in a formative state and yet the investment strategy calls for more commoditised solutions to simplify and therefore broaden adoption. Hence, progress will depend on new infrastructure that meets common and simple needs, as well more complex solutions for priority research communities and advanced users.

Australian Access Federation (AAF)

Easier, faster and simpler collaboration depends on the inter-working of multiple enterprise infrastructures. The AAF provides services critical to this inter-operation.

The AAF will support verifiable researcher identification, home based log-in (so-called Single Sign On), and community wide authorities and rights. The outcome will be simpler, more secure operation of all research community services.

The AAF is expected to support a wide range of research and higher education activities and therefore to be largely funded operationally on a user subscription model. Support is provided to assist uptake and develop services in (2008-2011).

Australian Research and Education Network (AREN)

Research activities lead to highly variable communication loads, often with high peak demands, distributed to unusual locations and with sporadic usage patterns. The AREN is configured to share basic capacity and relieve research projects from costs that might be associated with commercial supply of bandwidth for rare requirements.

AARNet operates the AREN and uses the core network infrastructure to provide commodity Internet access for its members. Revenue from these sources is expected to meet AARNet's operational costs. While discussion have included a second circuit to Perth or subsidising operational expenses, no new investment is in this plan.

The funds that have been allocated are expected to meet commitments from Government to fund the recurrent costs of links to Darwin from 2008.

AeRIC

The national e-Research infrastructure will face continual fragmentation due to urgency and independence. AeRIC will work to articulate a strategic view; to identify generic tools, practices and services; and match needs to available services.

The outcome will be a focus across the total investment on common needs and the ability to focus effort systematically on priority infrastructure development.

AeRIC will establish the National e-Research Architecture Taskforce (NEAT) to focus expertise and resources into teams targeting selected communities or new core services. The benefits will be the evaluation of systems, middleware and tools leading to the deployment of new services, and a more sustainable evolving e-Research fabric leading to increased e-Research adoption.

NEAT will identify, scope and support activities that include:

- projects to 'adopt, adapt and influence' global standards, components and middleware to create tools and services that meet identified discipline needs;
- the strengthening of key expertise groups associated with important existing or selected future middleware, user tools and global standards; and
- the hardening of new services in preparation for their on-going operation.

INDICATIVE FUNDING SUMMARY

Component	Arrangements and relationship to foundation principles	NCRIS	Others[2]
AeRIC NEAT	AeRIC will sustain the strategic motivation and promotion of e-Research. It will also provide the essential framework in which additional funding could be applied, through which additional parties could participate and by which the community can influence the direction of national e-Research infrastructure *(A,D)*. The Executive Director will lead component cohesion, undertake component reviews, and develop an annual business plan. AeRIC will establish NEAT to advise on the services and components that need to be made operational as part of the evolving national e-Research infrastructure. Activities sponsored by NEAT will build services that can better meet needs from discipline and NCRIS capabilities. The expenditure is included in the Australian National Data Service and Interoperation & Collaboration Infrastructure *(B,D)*.	2 (11)	0 (22)
National Data Management Infrastructure	A specific aim of the overall investment is to improve the management and use of research data and, in doing so, address the needs raised in the PMSEIC report on Data for Science. This component will implement federated meta-services to support cross-disciplinary data re-use *(A,B,C,D);* provide custodial and analysis services for some data collections *(C, D)*; and assist researchers and institutions and develop towards best practice *(A)*.	21[3]	24
National Computational Infrastructure	Capability computing is a national imperative and this component continues previous support in that area continuing the ANU hosted national merit capability computing services which operate with an internationally recognised quality of service *(C,D)*. Other investments will support specialist computing where such additional resource can be identified as a critical need to leverage the impact of NCRIS investments *(A,B,C,D)*.	26	32
Interoperation & Collaboration Infrastructure	Regional service providers are a key ingredient in e-Research infrastructure. This component builds a platform of inter-working national services across those regional providers and develops their outreach so that an increasing set of research resources can be connected. The arrangement will allow flexibility over time so that new resources can be targeted *(A,C)*.	20	12
AREN	AARNet provides Australia's national research backbone and the Australian Research and Education Network (AREN) service *(B,C,D)*. New significant investments have not been afforded priority for funding in this plan. NCRIS will, however, meet existing commitments to the AARNet infrastructure to sustain the Darwin services.	3	N/A
AAF	An organisational structure will be established to house the Australian Access Federation (AAF) and operate PKI and shibboleth services required for collaboration *(C,D)*. The use of those services is expected to include non-research applications and so a subsidised subscription model is proposed.	3	N/A
		75	68

[2] The contribution under "Others" would rise if other NCRIS funded capability investments and other research organisations co-fund PfC services or the states significantly co-invest..

[3] An implementation plan for ANDS has yet to emerge, these estimates are approximations only.

PART THREE RESEARCH INFRASTRUCTURE

Investment Criterion 1 An investment plan must result in excellent research infrastructure that addresses the national requirements of the relevant capability areas described in the NCRIS Roadmap

"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." [4]

While the consultations for PfC and the other NCRIS capabilities confirm the importance of these components, the importance of authentication and system-wide authorisation has become clearer as has the importance of tools as the primary means for broadening the adoption of e-Research methodologies. Therefore, support for *Tools* and *Authorisation* has also been planned as part of the investment.

The investment plan also takes into account the following sector characteristics:

- Institutions and federal and state agencies within the sector will spend internally on these same components and, in aggregate, spend far more than NCRIS.
- The inter-relatedness of components means that priority cannot be addressed by selecting some needs over others, for instance compute over data (or vice versa).
- The state of readiness of solutions, technologies, and associated products and services, is highly variable when viewed on a community or discipline basis.

[4] National Collaborative Research Infrastructure Strategy, DEST, February 2006.

At present, national general purpose e-Research infrastructure services are supplied by AARNet (budget circa $40M pa) and APAC (budget circa $20M pa). Outside of those organisations, the following can be noted:

- The APAC state partners spend about another $25M pa on additional activities, many of which relate to PfC.
- The Bureau of Meteorology and CSIRO jointly invest more than $10M pa in high performance computing (HPC) and data services within the High Performance Computing and Communication Centre, and invest in additional related infrastructure, which needs to inter-operate with investments by NCRIS.
- Data interests, such as Geoscience Australia, state and federal research agencies, and the humanities overall, need to be included within a common national e-Research infrastructure, in a co-ordinated approach to national data management as proposed to PMSEIC.
- NCRIS capabilities themselves will spend an estimated $10M pa on e-Research infrastructure within the various capabilities.
- Linkage Infrastructure Equipment and Facilities (LIEF) grants are in excess of $30M pa with some component in e-Research infrastructure, and universities themselves invest in e-Research infrastructure.

Fortunately several DEST-initiated research information infrastructure projects have explored and demonstrated e-Research capabilities over the last few years; so that much is known about the nature of the issues and the means for addressing them.

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

Finally, the aggregate infrastructure expenditure mentioned above exceeds $100M pa, and the aggregate ICT budget of research institutions is larger again. NCRIS investments need to add value to these investments and provide a framework in which they can inter-operate and be leveraged beyond their immediate intentions.

Vision

The vision laid out by the eRCC report has been consistently reinforced throughout the consultation process.

"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.

"The Australian community and economy will benefit from the advanced capability enabled by e-Research".

While some of these characteristics cannot be addressed by infrastructure development, the broad thrust certainly can.

Therefore: this Investment Plan builds on current activities to enhance, develop and deliver the infrastructure services needed to realise this e-Research vision.

Emergent Behaviours

National and international scale ICT systems are some of the most complex systems built and operated and have behaviours and consequences that are difficult to plan or even predict. The idea of 'emergent behaviour' has been developed to represent the fact that ICT development leads to changing social and economic behaviour because of complex interactions of multiple developments rather than because of individual technologies or functions. For instance, mobile phones have changed the way social interactions are planned and organised, producing a more dynamic and fluid day-to-day life style, which was not foreseen when the technology was developed.

An interesting manner in which to set goals in e-Research is to describe desirable emergent behaviours and develop activities that can lead to those behaviours by deploying ICT technological capabilities in infrastructure development:

Area	Emergent behaviour
Data	All research data is appropriately curated and retained; and co-ordinated services support easy location, access to, and analysis of, that data.
Grid	'Problem oriented' virtual infrastructures are routinely and easily constructed from any number of sensor, instrument, compute, data or visualisation resource, located anywhere.
Support Skills	A cohort of experts co-operatively operates and extends the infrastructure and readily assists users solve research problems.
HPC	An inter-operating infrastructure of peak, shoulder, institutional and departmental resources matches capabilities to needs and priorities.
Networking	An inclusive, high speed low latency network connects all researchers to all research resources, including international sites, using a non-volume based charging model.
Authorisation	Researchers control who can do what with their resources through a common method based on a single login for each researcher.
Tools	Users are able to work with familiar tools despite accessing multiple and remote compute, data, analysis, sensor or instrument resources.

e-Research Architecture

While crystal ball gazing in ICT is notoriously difficult, a medium term view of the above can be refined into architectural implications for e-Research infrastructure.

Possible Achievements in the Medium Term

Routine retention of research data is in place at custodial institutions, shared use of high capacity regional data centres is the norm, 'Google like' collection meta-services operate across all published Australian primary research data.

Multiple capability class computing facilities and associated infrastructure is in place, as are investments for regional shoulder and capacity class systems.

Commoditisation of 'ensemble style' compute capacity is achieved so that common user tools can run anywhere.

Co-operating centres of expertise in e-Research technologies, and co-operating centres of e-Research infrastructure delivery, operate and interact globally.

Many 'problem specific' infrastructures are in every day use, integrating multi-state resources.

An AAF is in routine use across the majority of research activities.

Campus infrastructure is harmonised with a functional AREN with appropriate reach and bandwidth.

Associated policy frameworks and best practice standards are published.

Longer Term Goals

Networking and access control will become invisible to research activities, with dedicated virtual networks supplied on demand, connecting large collaborations of researchers, and allowing data and analysis to be shared within high quality virtual presence delivered from theatres to desktops.

Exemplar information services will develop to provide access to knowledge through browser like functionality that front-ends advanced analysis tools and the automatic access and aggregation of appropriate data to answer user questions and provide 'expert advice'.

System level science and international research collaborations will grow to draw on the interaction of multiple tools and data sources from across disciplines to address problems at a national and global scale; created and continuously evolving by re-configuring common underlying services.

National e-Research Architecture

Data will trend towards larger data centres located on a national, regional or sub-regional basis, because raw retention is a 'lowest common denominator problem' with commodity solutions, economies of scale, and little if any business differentiation; and the need for co-location of computing and the delivery of ultra-high bandwidth integration will drive costs up in smaller sites.

High end computing will trend towards locating with large scale data, to support timely analysis over those holdings, and to support the information products that will become the most effective, and sometimes the only effective, means of providing access for many users.

Therefore the e-Research architecture will be a network of computing and data resources of various scales, coalescing over time; so that private computing exists with private data; as the data scales up, the computing will scale up; and as the data is shared, the computing will also be shared.

A second tier of large scale data sources will include aggregators of highly distributed data, large scale in-silico research facilities, and high volume image generators in physical research facilities.

As a consequence, the 'network' will separate into two layers, as the 'system-like' demand diverges from the 'human-like' demand. The human layer will be largely web based from researcher to researchers and services. The high performance infrastructure layer will be system to system and able to divert bandwidth to extreme scales of burst transactions, commit multi-point bandwidth for sustained periods to high levels of quality of service (QoS), and interact with transaction schedulers.

The research network will also become geographically diverse to better match the data demand, to reach from researcher to service, and to match the capacity and QoS needed system to system.

Above this foundation infrastructure, a distributed layer of standardised middleware services will grow to support workflows, and allow migration of analysis to data rather than of data to analysis.

The national architecture will then become one of inter-operating information services, where access to those information services implies access to data, computing and network resources, a middleware layer supporting workflows and migration of analysis to data, international agreements on content formats and meta data standards, and a shared expression of rights, roles and identities.

The architecture will require national services to assure integrity over the shared expression of rights, roles and identities by authoritatively aggregating and re-purposing appropriate information.

However, the unique mix of access and provenance that might be defined by the participants, sponsors and custodians for each data set leads to devolved content management. Hence the architecture must differentiate between content and systems; and a national policy framework and accreditation for shared services and service providers will be required.

Platforms for Collaboration (PfC)

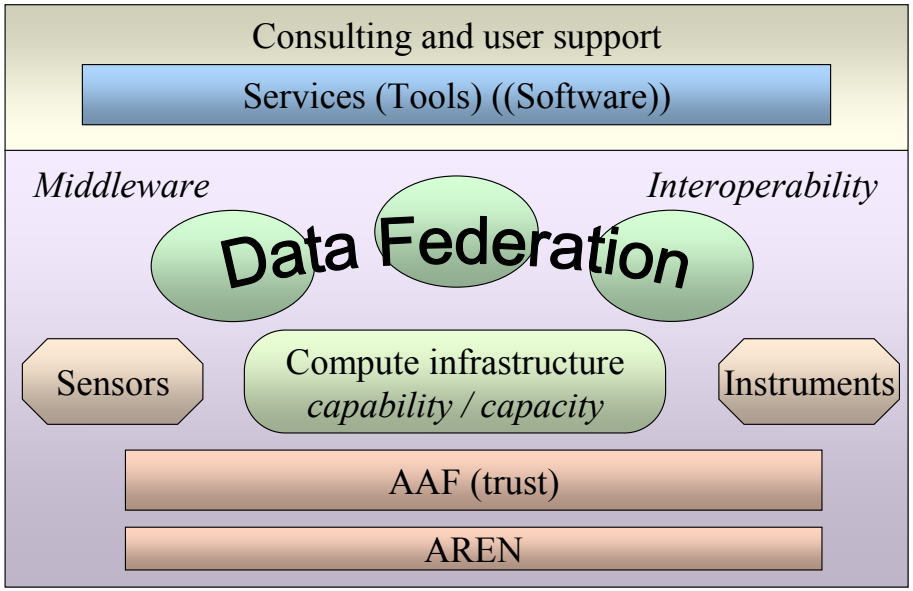
Given these over-arching directions, and noting the gathering dominance of data, the investment plan has developed around a view (depicted to the right) that – given fundamental capabilities in access (networking and authorisation services) and given the provision of compute infrastructure for processing, modelling and analysis, and given data gathering from sensors and instruments – the broader collaborative e-Research requirement can be best described as needs around data: for data access, data sharing, data analysis and data re-use.

It is in the communal use of data within and across multiple disciplines that e-Research opens the most valuable doors. From an infrastructure point of view, this requires a large variety of tools commonly in use across widely different disciplines to all be able to access and use the same data, and for researchers to be easily able to control who can do what with which elements of their data.

The NCRIS consultations have also strongly endorsed the view that, beyond access services for data, access by researchers to suitable tools and user support is a critical factor in enabling e-Research. Essentially, effective tools and information services are the only means by which e-Research can be made feasible for most researchers and of value beyond the current cohort of the "user-builders" of cyber-infrastructure.

Consequently, the investment plan focuses on a number of components of the overall situation:

- the continued provisioning of the AREN and its use and access policies;
- the provisioning of trust services that enable collaborative pan-organisational access control (AAF);
- the continuation and development of compute infrastructure, especially relating to capabilities beyond the reach of departmental or even institutional investment levels and which make sense to share and re-allocate over time (NCI);
- the creation of the Australian National Data Service (ANDS) that can provide key capabilities, including federated services, outreach activities and stewardship of some classes of data, as well as source the expertise and provide the leadership needed to crystallise data capabilities within research organisations;
- the provisioning of a national grid-like inter-operation environment that brings relevant tools with associated data and compute capabilities to researchers (Interoperation and Collaboration Infrastructure – ICI); and
- support for nationally integrated user support operating across all services and all service providers.



A set of basic capability components were developed with various communities, then a number of alternative arrangements were considered. The final framing of the investment has focussed around four components:

- a component on data infrastructure, to establish a national data collections service and to provide the expertise and outreach activity needed by the sector;
- a component on computational infrastructure, in the form of a peak general purpose computational system and additional classes of systems;
- a component on inter-operation and collaboration services, to support system inter-working, builders' tools, and user services, to all relevant resources; and
- a component to advance and harden the next generation of middleware needed for e-Research in selected disciplines and priority research areas, implemented across the other components.

A small component of funding will also be applied to support infrastructure in the AREN and the AAF, as well as supporting AeRIC.

Expertise

One of the most important issues for an effective e-Research infrastructure concerns the development and provision of appropriate expertise.

PfC itself and nearly all the other investments in e-Research infrastructure across the NCRIS capabilities will increase demand for expertise.

The investment plans 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 as follows:

- Expertise development and access can be enhanced by building groups of specialists rather than relying on unrelated individuals who can easily become overloaded and unresponsive.
- Different expertise and different levels of expertise are required during different stages of a community's 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 and that consensus is largely driven by practical experience.

So e-Research depends on standards; and standards development is always a long iterative processes. This means that researchers will necessarily undertake bespoke software development in order to continue their research while their communities converge on discipline standards.

Further difficulties then arise as reliance is placed on such software leading to a need for improved software engineering and particularly software productisation expertise.

The time required to develop highly generic standards, such as for general data management or job description, are even longer so that discipline specific middleware and tools can be expected to proliferate for some time to come.

Many of these issues were identified by the eRCC and the basic perspective developed by that committee remains valid and is reflected here.

A general solution to expertise is beyond the budget of PfC and is a broadly based education campaign is out of scope from an infrastructure investment point of view.

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

Therefore an approach to expertise development, its internationalisation and the means to access that expertise needs to be considered in the AeRIC business plan.

Specifically, each component should identify expertise related to system or data inter-operability, build expert groups at an international level of quality, support access to that expertise in an advisory mode, contribute experts to development projects on an agreed basis, and demonstrate the ability to participate in and influence global development of standards, components and middleware in relevant areas.

The goal will be to identify and develop pools of national expertise in important aspects of the middleware and service offerings created by PfC, particularly in those relating to inter-operability.

The Data Management Infrastructure Component

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

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

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

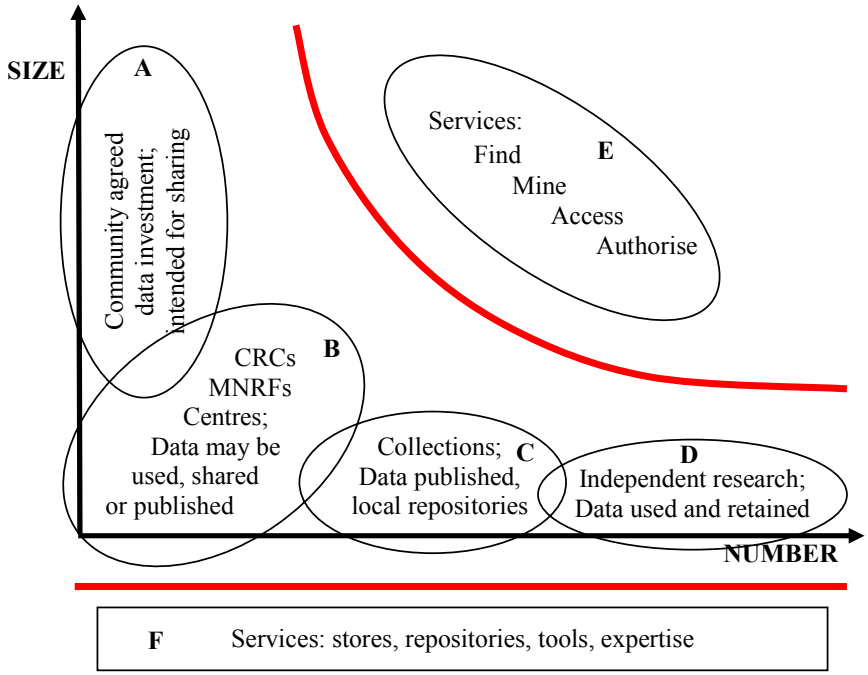
To create momentum in that direction, the component will implement a strategy to help experts and the builders and users of curation, access and analysis tools to work together to improve community data management practice — to 'keep data better'. It will provide:

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

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

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

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



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

(A) Several communities capture and share data of general value, where large datasets are generated or gathered as the result of significant investments; such as in astronomy, high energy physics, earth observations and bioinformatics. The size of these data sets tends to be very large, typically in terms of tens to hundreds of terabytes and sometimes petabytes, with high growth rates.

(B) More frequently, organisations, such as Co-operative Research Centres (CRCs) and Major National Research Facilities (MNRFs), have data gathering, generation, curation and publication as part of their mission; or use data for internal purposes, the results of which are published as an information service; or they provide research support services that generate data. The data often is privately held and only processed information is published. The size of these data sets tends to be in the many gigabytes to terabyte range.

(C) Even more frequently, many research organisations, departments, teams, and even individuals have established data collections the contents of which are intended for publication and access. These publicly available data sets (donated by researchers) are often kept in institutional or personal archives or web sites. The size of these data sets tends to be in the multiple gigabyte range.

(D) Nearly all researchers generate or store data on the desktop, much of which is only ever intended for individual use and which is inaccessible to other researchers, visible only by publication of derivative research results. Most of this raw data is intrinsically hard to re-use and could only be accessed if it was moved off the desktop. The size of such data sets is highly variable.

(E) Some organisations provide access to research data, either within disciplines or more broadly, and may or may not hold the primary data themselves. These organisations provide catalogues, abstracts or thumbnails, and can search and possibly mine data over which they hold indexes. These sources can only easily provide access to public data. The aggregate size of data accessible in this way would be in the range of terabytes to petabytes.

(F) A final class of mission relates to the retention, curation, access and analysis of data sets. Generic service providers necessarily operate only on public data, although some specialised services could operate on controlled access data. The co-location of data sets for data mining and linking is often part of the value of these services. Many institutions might operate such a service on their own behalf for their retained private data. The size of repositories tends to be in the range of hundreds of terabytes to many petabytes, and exabytes would be ultimately conceivable in some settings.

As the discussion suggests, data is everywhere, and the aggregate investment across all of the missions and user communities is large and rising quickly.

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

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

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

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

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

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

Institutions should:	Researchers should:	NCRIS 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; and ▪ provide secure systems for holding data and for granting access to that data.	▪ 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; and ▪ maintain confidentiality where it exists.	▪ provide national implementations of the federated services described under categories E and F; ▪ build expertise and provide outreach services that can assist others, including 'training the trainers'; ▪ broker solutions for collections and researcher needs and support identified collections to fast track e-Research development where appropriate; and ▪ ensure promulgation and use of agreed legal frameworks and templates governing access to data.

Implementation Strategy

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

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

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

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

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

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

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

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

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

AeRIC should ensure that:

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

ANDS should be empowered to:

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

Federated Services

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

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

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

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

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

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

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

Outreach Services

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

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

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

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

Stewardship Services

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

This activity includes three main components:

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

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

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

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

Next Steps

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

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

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

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

Stewardship

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

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

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

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

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

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

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

Federated Services

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

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

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

ANDS will support the installation and operation of servers and the configuration and maintenance of the services when operational.

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

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

Outreach

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

Some support will be provided in every region.

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

A significant portion of this effort may initially be related to the needs of collections identified for stewardship by ANDS.

The Computational Infrastructure Component

This component will bring together co-investors in major computational infrastructure and related service providers and work towards the vision:

> *An inter-operating infrastructure of peak, shoulder, institutional and departmental resources matches capabilities to needs and priorities.*

The component will make investments in computational infrastructure that support services which can best be funded and justified at the national level. A strong objective is to ensure the computational requirements of Australia's priority research can be met, especially through the provision of peak and high end computing and associated mass data storage systems.

The component will provide support on a merit basis and to the more purpose-specific computing capabilities needed by other NCRIS investments.

The capabilities will be extensible and allow third parties to co-invest for dedicated capacity or to 'purchase' capacity dedicated to discipline or problem specific services. The intention is to maximise options for economies of scale, to build-on rather than compete for expertise, and to provide a number of managed environments able to support computing capacity and services needed by other NCRIS investments.

The APAC National Facility has operated over the last six years supporting the high end, peak or capability class of user, with a quality of service and benefit to research that has been regularly reviewed. Each review, including the most recent in 2006, has praised the facility, its management and level of researcher support and commented that it provides an exemplary contribution to Australian national research infrastructure.

Countries comparable to Australia are planning the installation of additional major computing infrastructure to support leading-edge research. At the high end, the US National Science Foundation is envisaging a 'petascale computing' environment that would see at least one system with a peak performance in the 1-10 petaflops range by 2010, and the world's first petascale system is likely to be installed before that date.

Several important comments on HPC were made during the APAC review:

- The peak facility should be retained at or near the historical level.
- A more frequent purchase rate is needed to improve the return in Tflops delivered against dollars, assuming 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 resources are likely to be needed to allow for the broader clientele envisaged under NCRIS.

Estimates of likely demand suggest a ten-fold increase in compute and data storage capacity every 3-4 years. A reasonable but broadly stated goal would be to provide Australian researchers with access to at least one system in the 100-200 teraflops scale by 2009-10 and a petascale system by 2012-13.

It is also the case, that some of the computational needs of other NCRIS investments may be better met by providing shoulder or application specific systems. Examples of areas that could be provide a national service include:

- bioinformatics;
- geo-sciences;
- earth system sciences; and
- analysis support in characterisation, particularly at the synchrotron.

The access to these system resources needs to occur within an overall merit process.

Finally on expertise, the current trends in technology are fundamentally changing the nature of high end computing by introducing far greater levels of explicit parallelism in the processor. Over the last two decades, technology has introduced such parallelism implicitly by 'widening' a single processor to be capable of many more simultaneous operations but leaving these to automatic allocation within advanced code compiling technologies. The advent of multi-threading made processor parallelism more visible to programmers and the advent of multi-core can require more even explicit treatment of parallelism by programmers.

Therefore the availability of expertise to assist researchers adapt codes to these new architectures and to encourage a growing community of computational researchers is a small but vital component of an overall investment in high end computing.

The cohort of expert managers of high end facilities already co-operate on tenders, evaluations and purchase arrangements for each major purchase. Expertise developed by this investment should continue to be available to assist develop other HPC centres.

Implementation Strategy

The investment strategy for this component should build on the foundations that exist within APAC.

Therefore the implementation should provide for the following within the NCI Project arrangement:

- a Governing Committee with a Director charged with the ongoing development of an integrative framework for aligning significant compute infrastructure requirements and investments;
- a national peak capability service commensurate with international practice focussed on advanced services for the cohort of expert users with research goals that depend on such a capability;
- a national computational fabric complementing that peak capability with several application specific systems, supporting NCRIS priority research;
- a community based merit allocation system to consider system priorities and allocate resources against research community needs; and
- support activities to assist application development and porting and to develop a stronger network of computational researchers.

Due to timing considerations, the funding and the implementation arrangements should be initially based on the current arrangements applying to Systemic Infrastructure Initiative (SII) funding directed to the APAC National Facility. ANU should be invited to continue as the managing agent for the NCI Project.

Consideration should be given to changes that might be appropriate within the context of the establishment of AeRIC in light of its national role, and the governance and management arrangements adopted by the other PfC components.

AeRIC should ensure that:

- interim arrangements are in place by July 2007;
- the role and appointment processes of the NCI Project Steering Committee and any changes to arrangements are agreed and implemented by December 2007;
- NCRIS agreements are completed in consultation with participants and that continuing funds are provided under the revised agreement; and
- the process and decisions relating to shoulder systems are referred to AeRIC.

The NCRIS agreement should empower the NCI Project Steering Committee to :

- manage NCRIS resources provided within the terms of the business plan;
- ensure that third parties are able to make arrangements to support extensions of these systems for the use of specific research groups or organisations;
- undertake regular reviews of the needs of key user groups, determine the most appropriate upgrades to systems to meet these needs and recommend the allocation of funds to support those systems;
- ensure a merit allocation scheme is extended to grant resources on all facilities funded as part of the component through a cooperative decision process that matches user needs and available systems and resources;
- develop and sustain an outreach activity that assists computational researchers improve their use of high end computing;
- appoint the Director;
- report to AeRIC on the implementation of the business plan on a regular basis; and
- provide a progress report and propose a revised business plan on an annual basis.

The overall financial commitments, as currently envisaged, are as follows:

	Cash ($M)		In-kind ($M)	Total ($M)
	NCRIS	Members	Members	(av.)
Peak systems	18-20	10	10	39
Shoulder systems	2-4	6-8	5-7	16
CT&T	2	0	2	4
Management & outreach	2	0	0	2
TOTALS	24-28	16-18	17-19	61

However, the non-NCRIS commitment to any of the facilities could increase, opportunities for co-investment in the systems will exist during and after purchase decisions, and the split between peak and shoulder systems may need to be reviewed within a revised business plan, as the best choices are technology and use dependent.

National Peak Facility

It would be possible to consider opening the hosting function of the National Facility to contenders other than ANU; however, this is unlikely to lead to any increase in funding for the share available to researchers by merit or through priority assignment, and might dilute expertise and service quality and therefore is not recommended.

The current strategy of acquiring a peak system every 3-4 years should be revised to increase the frequency of increments to performance, either by overlapping slightly smaller systems, achieving upgrade options or through lease style replacement options.

The specific financial planning for the peak system is highly detailed and depends on expected purchase points, cash management between NCRIS and the host agency, and costs associated with considerations around operational overlap.

The NCRIS funds identified in the financial outline would be expected to yield a system in the 40-50 ranking on the TOP 500 list at the time of purchase.

The national capability facility should also provide systems, specialised software and associated expertise to support data intensive applications. The data may be generated from the computational systems or be part of a data collection hosted by the ANDS.

The facility should also provide and support software consistent with having unique and specialised system. This software would include application packages, support tools, compilers, databases and file manipulation tools and parallel programming libraries, performance tools and debuggers. The software would need to be regularly updated and associated tools installed as improvements are sought by users.

National Computational Fabric

The national computational fabric should aim to provide the underlying computing capability that best supports the key disciplines and priority research of NCRIS investments.

The details of this investment should include co-investment in systems and system support operated by regional service providers.

NCRIS investment in a national computational fabric will trigger the development of more significant computational capabilities in regions and make available a share of such systems for allocation under a common compute infrastructure merit allocation system. It is expected the overall systems should support the capabilities and priority research areas identified in the business plan for that system.

This investment will make available computational resources to researchers at a national level.

Arrangements should permit third parties to expand these systems through a separate contract with the regional service provider operating the system, where the proposed use aligns with the research support intentions of NCRIS funding. Such extra capacity could be allocated on whatever basis is agreed between the service provider and its customer.

This strategy will lead to several sites being funded to provide specialised systems and skills relevant to priority research needs. It will also mean that the peak computing system can be focussed more on ‘capability’ computing and a more diverse range of systems and support will be available to respond to evolving research requirements.

Computational Tools and Techniques

An activity should be established to improve the software development environments for users of advanced computational systems.

There will be an increasing need to provide researchers with problem-solving environments which expose the methods to solve the problem and hide the complexities of the systems.

Other areas that will need special support relate to efficient coding for multi-core chips and performance accelerators. The foreshadowed changes in system architectures will have a major impact on all forms of advanced computing.

Other areas for development are likely to relate to applications software. It can be expected that some activities will contribute to international efforts by adapting codes for the local computing requirements.

Outreach

An outreach activity should be established to engage the broader research community in gaining an appreciation of the value and practice of high end computing.

It should market advanced computing to researchers and research managers and provide opportunities for networking that can lead to cooperation and skill enhancement through collaborative projects.

The activity should find ways to reach out to the whole computational community.

The Interoperation and Collaboration Infrastructure (ICI) Component

This component provides the operational and the development arm of the integrative infrastructure underpinning PfC.

The operational activity brings together providers of shared compute and data services to operate an extended "Australian National Grid", working towards the vision:

> *'Problem oriented' virtual infrastructures are routinely constructed from sensor, instrument, compute, data and visualisation resources.'*

The kinds of future functions envisaged are typically described in terms such as:

- on-demand, ubiquitous access to remote computers, digital repositories, scientific instruments and most recently sensor networks;
- virtual environments and on-line collaborative interaction;
- seamless resource sharing;
- cooperative working; and
- coordinated problem solving within dynamic virtual organisations.

The ICI component should install, configure and operate servers and services that allow grid like workflows to use resources anywhere in the country, subject to suitable authorisation. This kind of capability is essential for research in disciplines such as high energy physics and climatology, as well as AuScope, e-Marine Information Infrastructure (eMII) and Nanostructural Analysis Network Organisation (NANO).

As a general observation, many disciplines are developing standards and functionality that meet their own needs, and have low tolerance for the kinds of timescale that would be needed to agree generic middleware or user interfaces. Australia therefore faces the problem of playing catch-up as global disciplines have sufficient funding to develop their own infrastructure components and effectively 'go their own way'.

Australia could create its own unique national grid flavour or adopt one of the international flavours of grids over the others. However, neither option is particularly viable given the scale of investment required in the first case and the unsupportable discrimination that would arise at a disciplinary level in the second case.

There is however some value in adopting and supporting a limited set of grid flavours, especially as international efforts are working towards the inter-operation of a number of major grid flavours. In particular, projects exist to enhance inter-operation between globus, glite and unicore grid middleware.

The last point aligns with some very clear feedback from consultations in this area, which was that it would be best to focus on providing core services which are certain to be of value and to do them well.

Five such key contributions can be identified at a cyber-infrastructure operational level, which application oriented workflows and user oriented tools and portals need, and which can be operated as services in common.

Therefore the proposed initial activities of the Interoperation and Collaboration Services component will need to include:

- a middleware operations function that can support several middleware and portal flavours operating over an integrated resource allocation and authorisation system;
- a job submission fabric that allows the automatic unattended submission of jobs and workflows on any system within the fabric;
- a data movement fabric that allows the automatic unattended transfer of data including effective large bulk transfers between any systems within the fabric;
- a set of easy to use virtual organisation services, including AAF integrated collaboration workspaces;
- the day-to-day operations management for the above services; and
- a small activity responding to requests for help and tacking requests through to expertise groups and services and their response.

Investment into these services will need to include the provision of systems and servers, software and software environments, and the management means to ensure they are reliably configured and regularly tested for compliance and operational status.

The intention is that as additional requirements emerge from NCRIS capabilities, and as responding PfC services become clearer, those services will migrate into this 'operational arm'.

As an example, participants in ICI might supply and operate federated services and data retention services specified by ANDS, or the selected suppliers of those services might join ICI to ensure reliable national inter-operation.

The component needs to build on SII investments to date, not just in the APAC grid but also in other knowledge management projects.

In moving forward, two classes of participants are envisaged, namely Partners and Affiliates, and a broader set of technology options will need to be supported.

Partners should be service providers supporting compute, data, visualisation or collaboration services to researchers across multiple institutions. As partners, they will install and operate core services on their premises and commit to configuring their own systems and services to integrate with agreed protocols and standards.

A goal is to have at least one participant with partner status in every region.

Potential partners include:

ANU, ac3, QCiF, SAPAC/eRSA, TPAC, VPAC, iVEC

Affiliates should be organisations that seek to allow access to their research resources through these services or to seek access to other resources through these services. Affiliates will generally not be the operators of the core infrastructure services.

Affiliates will need to agree to configure their systems to comply with agreed protocols and standards and the regional partner will provide assistance as required. That assistance may extend to co-managing servers on affiliate sites if that is required for reliable operation.

Potential affiliates include:

AA, AAF, ABS, ACcESS, ANSTO, AuScope, ATNF, BoM, CSIRO, eMMI, GA, IMB, NLA, Various Medical Institutes, Synchrotron

Many Universities

ICI will be led by a Director appointed by the participants, and will require a number of system administrators to be appointed at each partner site.

Partners will need to provide these staff with the necessary site privileges to configure and operate the servers and services required by the ICI activities.

However, operation can only be on a best effort basis as the evolving middleware environment often creates incompatibilities and failures that only become evident as users access unique combinations of features and resources.

Implementation Strategy

Therefore users will need to accept that there are risks related to functionality and availability at any point in time.

AeRIC should ensure that:

- an interim ICI Governing Committee is established to provide guidance on further development;
- expressions of interest are invited in the roles of contract agent and members in consultation with the interim governing committee;
- an NCRIS agreement and business plan based on the attached implementation plan is developed, in consultation with the interim governing committee and proposed participants; and
- the NCRIS agreement and business plan is finalised by December 2007.

The NCRIS agreement should empower ICI to:

- receive funds from AeRIC under the arrangements defined in the NCRIS agreement and against the activities described in this investment plan and the subsequent business plan;
- operate as a single co-operative activity across all members;
- develop a community of interest around inter-operation standards and services and engage in relevant international developments;
- develop and sustain the services for middleware support, job submission, data movement, and VO services identified within the business plan;
- establish and sustain an operation centre able to track and manage configuration requirements, identify grid performance, rectify faults and provide expertise to support infrastructure builders and operators in other programs;
- establish and sustain a help service which can direct user queries to the expertise most likely to be able to assist and track responses and resolutions;
- appoint the ICI Director;
- manage resources within the terms of the NCRIS agreement;
- report to AeRIC on the implementation of the business plan on a regular basis; and
- propose a revised business plan on an annual basis.

Initial ICI services

Middleware Hosting

The ICI will need to identify and support a small set of inter-operable middleware protocols (expected to include globus and glite) and a small set of inter-operable data access protocols (expected to include GridFTP, SRM and SRB) on all partner sites.

The ICI will also need to identify core application support services such as gridsphere and arrange for their suitable installations to be available on the ICI infrastructure.

Job Submission Fabric

The ICI will need to select standards and information models that allow authorised users to submit jobs on any partner's or affiliate's resources and will also need to standardise job submission semantics and protocols across the members' systems.

Data Movement Fabric

The ICI will need to select standards and information models that allows any appropriately authorised users to access, replicate, and relocate any data associated with any partner or affiliate resource. The ICI should be responsible for the reliability and speed of the data movement services.

Collaboration Services

The ICI will provide a range of user collaboration tools including AAF integrated wikis, shared workspaces, and email lists, and should support access grid systems at each partner site and as many affiliate sites as is practical.

Operations and User Support

The ICI will need to establish an operations centre that can assure the reliable inter-operation of the participating systems and sites. It will also need to deploy and manage servers, such as gateways and compute and data access servers and provide a web based virtual 'operations centre' for real time and historical information.

The ICI will need to provide a "User support" function as a contact point, a web and email presence, and a ticket issuing and tracking system.

Budget

ICI members will be regional service providers which support compute, data, visualisation and collaboration services to researchers relevant to NCRIS (and others) across multiple institutions. ICI members will need to demonstrate their ability to engage regional research institutions and communities, and undertake work to connect key research facilities in their region into the national e-Research infrastructure.

ICI is not to be a funding stream for internal ICT services or for high end computing or data systems located within operational units.

At the financial level, NCRIS will need to ensure that funds support specific staff and systems functionality related to nationally co-ordinated core data and computation services. Affiliates should use other resources to extend those to meet local needs.

A decision needs to be made on the eligible regions, noting that APAC allowed one per state and that perhaps this raises difficulty in the more populous and highly dispersed states (such as Queensland).

The business plan and NCRIS agreement will need to include:

- the arrangements and responsibilities for the ICI governing committee and the ICI director;
- appropriate service level agreements; and
- an initial schedule assigning responsibilities to proposed participants and the process by which that schedule can be varied over time.

The initial NCRIS contribution established for the activity has been set at $3.5M per annum with an expectation that up to a further $1.25M from NCRIS and a minimum of $2.5M per annum in kind from ICI members might be available for development, application, expertise and user support activities authorised by NEAT.

	NCRIS ($M)	Other ($M)	Total ($M)
Operational Services	14	0	14
Developmental (NEAT)	6	12	18
TOTAL	20	12	32

Foundation Infrastructure

The goals in this area involve the fundamental electronic and human interconnects that make collaboration possible, through the two emergent capabilities:

> *An inclusive, high speed low latency subscription-based network connects all researchers to all research resources, including international sites.*
>
> *Researchers control who can do what with their resources through a common method based on a single login for each researcher.*

The collaborative nature of e-Research and much of the activity that is envisaged under NCRIS will be significantly dependent on an ability to share data, systems and research infrastructure between research organisations and researchers, both within Australia and internationally.

At present, sharing data, systems and research infrastructure, in a secure way, between various universities and including research agencies presents a very difficult problem, both technically and from a policy perspective.

Therefore the two challenges contemplated in this area are as follows.

- to confirm the concept of the AREN and to implement steps to align the backbone, regional and campus network system elements with that vision; and
- to build a national access control system that supports effective collaboration between institutions and secures resources and access to known users.

These services, have a much broader use than research and, while mentioned here in terms of research objectives, will add value across the research and higher education sectors for many other uses.

DEST has supplied $88M from the Systemic Infrastructure Initiative (SII) to assist the establishment of the Australian Research and Education Network (AREN) and has recently provided $4.8M to assist establish the Australian Access Federation (AAF).

The proposition is that operational costs for these activities should be largely funded by the institutions gaining value from them, so that:

- communication and authorisation capabilities are focussed by their users;
- a co-operative rather than directive approach to infrastructure is enhanced; and
- further development is naturally sensitised to cost benefit.

A very large number of potential developments could be funded in these services. For instance:

- The AREN backbone could be extended with layered services to Perth, into other regional areas and offering higher levels of redundancy to all points of presence.
- Investments into significantly increased capacity could be made to support future requirements and to improve access to international facilities.
- Funding could be provided to harmonise campus infrastructure with the AREN, to deal with the ‘last mile’ problem in this context, namely that effective access to the AREN is limited by campus networks.
- A range of activities and actions could be undertaken (and funded) to harmonise policies between the national, regional and campus network components.
- A larger and more complex approach could be funded to attempt to implement the AAF as a component within other government authorisation initiatives.
- Funding could be provided to adapt a range of applications to integrate with the AAF.

Overall the funds available to PfC represent a reduction to about 40% of the fund level available within SII, on an annualised basis.

Consequently the existing AREN foundation is treated as adequate for the next four to five years, noting that further development can always be made on a co-operative or internally funded basis.

The risks in a ‘big bang’ approach to establishing a trust federation are substantial with little obvious gain over a more evolutionary approach. Therefore the development of the AAF is planned to take time and its integration with an increasing range of research support services will depend on co-operative funding by the beneficiaries.

The Australian Access Federation (AAF)

Recent technical advances provide a new solution to secure sharing of data and systems using a "Trust Federation". A Trust Federation is a combination of policies and technologies that allows individuals across many different organizations to securely share and analyse data within a trusted environment.

Some of the very significant advantages of a Trust Federation are that a researcher needs only one name and password (or its equivalent, such as a digital certificate) to access all appropriate resources at different institutions and that these credentials would be issued by the researcher's home institution.

Additionally, an institution will need only one agreement with the trust federation rather than 40 or 50 individual agreements. The federation can also help to reduce costs through the collaborative purchase and/or support of some tools and technologies.

Previous DEST funding on Trust Federations through the SII, particularly the MAMS (Meta Access Management System project based at Macquarie University) and the first phase of the e-Security Framework project (based at the University of Queensland) have provided a foundation for an Australian Trust Federation.

For instance, the MAMS test bed Trust Federation already has nine universities with over 600,000 members. The e-Security framework project has established a PKI infrastructure with inter-operation between different universities, has obtained "in principle" agreement from the US Higher Education and Federal Government PKIs for interconnection with the Australian federation, and has made significant headway with Microsoft and other vendors in updating their browsers to support the Australian PKI Infrastructure.

PKI and Shibboleth are different but complementary technologies for supporting a Trust Federation, and both are required.

PKI is typically used, as an authentication mechanism, for small numbers of users (100-1000s) in relatively high security contexts, but is also used for other collaborative activities, such as encryption and digital signatures.

Shibboleth is typically used for larger numbers of users (100,000-1,000,000s) in low to medium security contexts; although increasingly both technologies are used together as part of an overall "trust fabric").

The Australian Access Federation (AAF) will be established as a 24x7 operational entity to provide these basic services to the Australian research community.

The AAF will provide:

- accredited highly secure servers for key infrastructure functions;
- PKI and shibboleth services to the entire research community;
- services as required by the ICI infrastructure; and
- expertise in these areas to other participants within the infrastructure.

DEST have funded the development of the AAF with $4.8M from SII to facilitate the entry of Australian universities and research institutions into a trust federation. The AAF will undertake start up activities for the federation and will support the work that institutions must undertake to be able to fully participate.

This support will be provided through training, development and implementation of shared technical infrastructure, common approaches to support and procurement, and common approaches to vendors for system integration.

The aim is to significantly reduce the financial, technical and organisational barriers to collaboration across the sector. Work will be undertaken in three parts:

Part 1: The development of overarching governance and policies for AAF (these are independent of delivery technologies);

Part 2: The development of specific policies, technical implementation and rollout of PKI for AAF; and

Part 3: The development of specific policies, technical implementation and rollout of Shibboleth for AAF.

In order to fully participate in the AAF, each institution will also need to provide and maintain some infrastructure for identity and certificate management services.

An allocation of $3M of NCRIS funding ahs been provided to assist the early years of operation of the AAF; however, it is intended that the AAF would operate on a subscription basis. The intention is that any of these funds unused by the AAF will be applied through NEAT to develop AAF integrated services.

The Australian Research and Education Network (AREN)

An advanced high speed and reliable network and network services is a critical foundation component of eResearch Infrastructure.

The Australian Research and Education Network represents this component in that the AREN is a vision of education, research and government networks interconnected via a very high speed Australian Academic and Research Network - carried on AARNet3.

It is inevitable that some of the entities where NCRIS researchers are located will be beyond the AREN reach and hence some use of the commodity Internet will be unavoidable in addition to its use to support mobility.

However, as far as practicable, the AREN will interconnect researchers and major facilities (i.e. compute, data and instruments) via the multi-gigabit AARNet3 backbone with a preference for connecting major research facilities directly to the AARNet3 backbone or the equivalent AREN regional component.

The networking requirements of high end e-Researchers (based on examination of the NCRIS capabilities and analysis of traffic patterns and consultations with researchers by the US Office of Science) are likely to require the following capabilities:

- increased capacity;
- high network reliability;
- high-speed, highly reliable connectivity with international R&E institutions; and
- new network services to provide bandwidth guarantees.

The Office of Science analysis showed that 50% of the traffic was generated by the top 100 sites in the US with an expected ten fold increase over the next four years (without the production Large Hadron Collider traffic). The analysis also showed a significant traffic pattern change in that 72% of the hosts generating the top 1000 flows were parallel data movers resulting in peak flow rates decreasing whilst total traffic increased.

The network architecture to deliver these requirements will include:

- A high reliability national and international gigabit IP backbone supporting advanced features not as yet commonly available from the commercial telecommunication carriers (e.g. jumbo frames, IPv6, multicast).
- The capability to provision guaranteed very high speed switched and point to point circuits both permanently or for specific periods to support large, high-speed data flows and meet unknown bandwidth requirements by provisioning additional lambdas.
- Regional and institutional networks that provide dual site connectivity for reliability to the national IP backbone and the capability to provision switched and point to point very high speed circuit based links.

AARNet3 and its partner regional networks like VERNet and SABRENet are well advanced in meeting these requirements for universities and CSIRO due to major investments by DEST, some state governments, and members over the last three years. These investments have provided the Australian research and education community with rights over 10,000km+ of dark fibre for the next 10 to 15 years. However, the requirement for advanced end to end network services will require harmonising campus infrastructures with the AREN. An activity to assist institutions with a source of design and implementation advice and to harmonise national, regional and institutional network services will be developed and delivered under the expertise component of Application Services.

It is also recognised that there are a number of "gaps" in the AREN infrastructure and that some research entities do not have the redundancy to meet the anticipated reliability requirements. Unfortunately, with limited NCRIS PfC funding and other priorities, it is only possible in this investment plan to provide support to sustain the link to Darwin beyond 2008. It is possible that other programs such as the DCITA *Connect Australia* initiative will offer solutions in this area.

Work also needs to continue to interconnect the AARNet3 and regional networks within the AREN vision to other research agency and government networks where e-Researchers are located. However, the AREN policy relating to these "tail" connections is that they are the responsibility of the connecting entity with DEST research infrastructure funds focussing on national backbone and international links.

The programme activities should also assist the AREN by bringing more researchers into its scope, and the ICI and ANDS components can be expected to significantly increase research on-net traffic, assisting the value proposition of the AREN to institutions.

A further issue that will be addressed is the provision of a source of design and implementation advice that can harmonise campus infrastructures with the AREN.

PART FOUR OWNERSHIP AND MANAGEMENT

Investment Criterion 3 An investment plan must include a facility ownership and management structure that will result in the efficient and effective operation of the infrastructure

Governance

Because the pace of change in e-Research cyber-infrastructure is high; it is likely that change will be needed to the components over the investment timescale. In addition, additional funding for related activities may become available which may also require re-planning and re-aligning of activities. This suggests that the maximum flexibility be retained (commensurate with good management of the programs of activities within the investments).

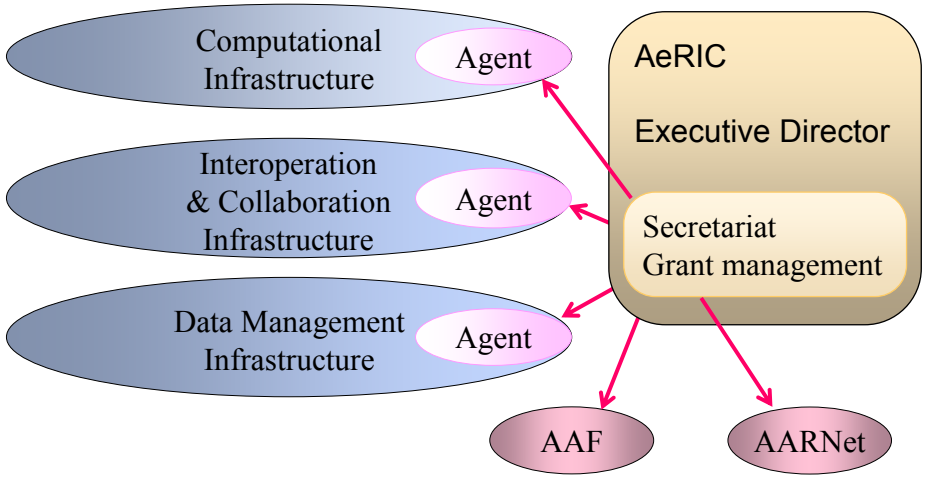
A high level of co-ordination effort is also implied if investments are to be brought together within the changing environment to achieve the emergent capabilities needed by e-Research. This suggests a single overall managing agent is needed.

On the other hand, the various activities have very different properties, and many of the relevant parties have interests well outside of the NCRIS framework and indeed outside of the research sphere. This suggests the need to separate activities so that like interests can be involved in areas that matter to them and organisations are not compelled to be involved in activities that distract them from their own missions.

The entities that may be involved are also likely to change over time, so that the contractual arrangements need to be 'light weight' and flexible. On the other hand, while the activities may include a few participants, many entities and research groups will come to depend on them, so that contractual obligation around service delivery and arrangements for community steering will need to be in place.

An overall structural framework that meets these various requirements is depicted to the right. In this framework

- The primary strategic governance for activities within PfC will be through a high level committee established by DEST, to be called the Australian e-Research Infrastructure Council (AeRIC);.
- Primary oversight will be through the AeRIC secretariat responsible for the implementation of the business plan approved annually by the NCRIS Committee.; and
- Financial control will be via contracts between DEST and the lead agencies and participants associated with component activities.



The result is a fast track and flexible start-up for PfC, as the secretariat can be established quickly. Activities which need continuation, such as the APAC National Facility, can also be easily continued under a suitable grant. The arrangement also allows for incorporation at a later date if desired.

The Australian e-Research Infrastructure Council (AeRIC)

The Australian e-Research Infrastructure Council (AeRIC) will be charged with setting direction around the use of, and driving the uptake of, the PfC infrastructure.

The goal of AeRIC will be to provide an holistic oversight of the activities contributing to the development of e-Research capability.

AeRIC will have an independent Chair, an Executive Director, and its membership will include a nominee of DEST (as a Principal Adviser).

DEST will appoint the Executive Director who will be an ex-officio member of AeRIC.

The remaining members would be appointed for their expertise in the provision of e-Research capabilities or for their expertise in applications of e-Research capabilities.

The AeRIC will:

- approve strategic directions for new components or for variations to existing components under the PfC banner;
- convene an e-Research community forum to provide advice;
- engage collaboratively with DEST in relation to its deliberations and determinations, including through the Principal Adviser DEST who will have a seat on the AeRIC;
- approve the annual business plan prepared by the Executive Director and the activities operating under the PfC banner; and
- make recommendations to DEST and the NCRIS committee arising from the business plan, including for allocation of funding and programme variations.

AeRIC would provide a formal progress report to DEST annually but would be able to make recommendations to DEST at any time in response to changing needs or developments in e-Research.

The AeRIC secretariat budget and staff levels will form part of the PfC annual business plan and will require AeRIC approval. The Executive Director will have the necessary authorities to expend those funds, including to engage such additional staff as may be required to administer the activities nominated in the business plan.

The secretariat will focus on two principal business activities, namely programme development and the management of the funding contracts related to the PfC components.

NEAT

AeRIC will establish a National e-Research Architecture Taskforce (NEAT) to:

- work with PfC components, other NCRIS capabilities and relevant discipline areas to develop a timeline for infrastructure development;
- develop activities for the trial and hardening of new services noting that where activities are proposed over multiple years, annual reviews and justifications will be required; and
- recommend projects, participants and funding arrangements for these projects to AeRIC.

NEAT will be chaired by the AeRIC Executive Director during its start up phase. An independent chair may be subsequently appointed by AeRIC.

An allocation of $10M of NCRIS funding will be provided to NEAT which is expected to be trebled to a total effort level of $30M over the four years.

The ‘hot spot’ for activities is expected to be the development of tools and services related to the common interests of ANDS and ICI.

Component Governance

While the establishment of multiple components does allow different participation and conditions to be applied in each implementation activity, it also leads to multiple agreements. In general, the number of agreements needs to be as few as is possible.

Implementation component members may also need to satisfy a qualification hurdle and accept the obligations inherent in the implementation of the component.

The component business plan will be approved by AeRIC.

The general arrangements around a component will include the following:

- A component director will be appointed, to be at least a half time position.
- The role of the component director will be set out in the business plan; one element of which will be to develop a revised business plan each year.
- The component director, and other representatives as set out in the business plan will form a governing committee for each component.
- The component governing committee will be tasked with oversight of the component and endorsement of changes to the component business plan.

Apart from the usual issues to be covered in a partnership style agreement, each component agreement will also address the following items:

- the categories of members and the associated rights and obligations;
- the process for identification and admission of new members and the processes for termination of membership (voluntary and compulsory); and
- the identification of matters which may be determined as part of component activities (such as merit allocation of resources), by the component director (such as resource substitution); or by the component governance committee (such as commencing a new activity, or re-assigning resource between activities).

Agreements should be arranged so that members have the right to terminate their participation without prejudice, after meeting any outstanding commitments, on three months notice.

Lead agencies must have the ability to terminate an agreement with any component member on three months notice in accordance with agreed processes; for any of the reasons identified as part of the business plan, such as performance failure or financial variations to the component imposed by AeRIC.

Each component plan will need to address the following items:

- the overall goals of the component and the objectives set for the year in question;
- the key performance indicators that will be used to assess the component's effectiveness and contribution;
- a summary of resources available to the component, both funded by NCRIS and contributed from other sources;
- a project plan and associated milestones for each activity identified within the business plan; and
- an organisational chart identifying the personnel and their roles and authorities within the component's activities.

Each activity within a component should be required to provide a quarterly report summarising achievements and notifying any variations to resource levels or milestones to the component director.

Where funding is provided to component members, it should generally be provided quarterly in arrears following submission of the relevant summary. Agreements should allow further funding to be withheld where the effort level required by an activity has not been provided, until such time as the effort contribution is made good.

Funding for major acquisitions will need to be arranged separately.

Member contributions need to be managed on a best effort basis, so that payment while in arrears is not necessarily milestone based. However, significant failure to perform as judged by the component governing committee should result in the replacement of the service provider and the reallocation of the services to alternative providers.

Where a component governing committee or AeRIC determines that project failure has occurred, activities may be terminated or curtailed, and funding reduced if substitute activities cannot be agreed.

Component members should be able to request the substitution of staff, and variations to resource levels within and across projects within a component, in order to assist their maximum contribution to the overall investment plan.

Ownership

Computational Infrastructure	National Facility	ANU will hold the systems in the National Facility purchased with NCRIS funds on behalf of and for the use of NCRIS community.
	Computational Tools and Techniques	No physical assets are envisaged under this activity. Intellectual Property generated by these projects will rest with the parties participating in the projects. Improvements to any pre-existing IP will rest with the owner of that IP.
	Other Facilities	The organisation hosting and responsible for the facility will own the assets and will provide services to users determined by the Merit Allocation Committee commensurate with the contributions from NCRIS.
Data Management Infrastructure	Stewardship Services	The intention is to purchase services, however should joint investment in assets arise, those underlying assets would be owned by the third party investor and PfC would gain a commitment on a service, access or share basis. Intellectual property related to the operation of those assets and improvements to them will belong to the third party.
	Outreach	No assets are envisaged under this activity. Intellectual property relating to improved standards and processes will be placed in the public domain third party participants would retain any contributed IP and any improvement to that IP.
	Technical Services	The intention is to purchase services; however, should joint investment in assets arise, those underlying assets would be owned by the third party investor and PfC would gain a commitment on a service, access or share basis. Intellectual property related to the operation of those assets and improvements to them will belong to the third party.
Interoperation and Collaboration Infrastructure	Developmental aspect	No physical assets are envisaged under this activity. Intellectual Property generated by these projects will rest with the parties participating in the projects. Improvements to any pre-existing IP will rest with the owner of that IP.
	Operational aspect	All assets will be owned by the members and affiliates and contributed on an in-kind basis. Intellectual property generated by the activity will be placed in the public domain.
Foundational Infrastructure	AREN	In view of the recent considerable Australian Government investment in the AREN, the only funded activity in this plan related to the AREN concerns the operating costs of a link to Darwin.
	AAF	Funding provided to the AAF will be to defray operating costs and will not be in the form on an NCRIS agreement. Any assets and the disposition of IP will be determined by the implementation of the AAF.

PART FIVE ACCESS AND PRICING

Investment Criterion 2 An investment plan must result in research infrastructure that is accessible by researchers on the basis of merit, at reasonable prices, and that encourages collaboration in research

(A) Some services supported by PfC will be open access to all researchers (such as AAF, AREN, operational ICI, technical services in ANDS), some services may be public access.

(B) Access to some systems and services supported by PfC will be by merit determined through a merit allocation process that operates independently of the organisations providing the services and is open to any researcher from the classes of organisations supported by NCRIS. A separate merit process will apply in each component.

Computational Infrastructure	National Facility (B)	Organisation may also purchase a portion of the systems for their priority use at full operating cost. The cost to research organisations will be the annual cost of the facility in that year factored by the share made available. The cost of access to commercial users should be set at a higher rate.
	Computational Tools and Techniques (B)	An open call will be made for project proposals that can assist users better utilise any computational resource funded by PfC. The governance committee will establish an open review and decision making process. A factor in merit for such activities will be the resources made available from other parties to the project, and the demonstrated ability to pursue the project as a single activity combining resources from multiple sources.
	Other Facilities (B)	The NCI Project Steering Committee will establish a process by which proposals can be formed around the establishment of application specific computing installations. Proposals will only proceed to implementation with AeRIC endorsement. Access to such systems will be on the same terms and through processes as applied at the National Facility.
Data Management Infrastructure	Stewardship Services (B)	ANDS may also broker arrangements between providers and collections, which may require full cost funding by the collection. Collections may channel funds through ANDS for that purpose – however, this is to be determined.
	Outreach (B)	ANDS may also manage outreach services funded by participants on a pass through basis.
	Technical Services (A)	NEAT will determine the developmental activities that may be undertaken in this category; participation will be by invitation.
Interoperation and Collaboration Infrastructure	Developmental aspect (B)	NEAT will determine the developmental activities that may be undertaken in this category; participation will be by invitation.
	Operational aspect (A)	Some services may be restricted to partners and affiliates for technical feasibility or reliability reasons.
Foundational Infrastructure	AREN (A)	In view of the recent considerable Australian Government investment in the AREN, the only funded activity in this plan related to the AREN concerns the operating costs of a link to Darwin.
	AAF (A)	Funding provided to the AAF will be to defray operating costs and will not be in the form on an NCRIS agreement. Any assets and the disposition of IP will be determined by the implementation of the AAF.

PART SIX IMPLEMENTATION STRATEGY AND BUSINESS CASE

Investment Criterion 4 An investment plan must include an implementation strategy and business case that will result in the efficient implementation and effective ongoing financial management of the infrastructure

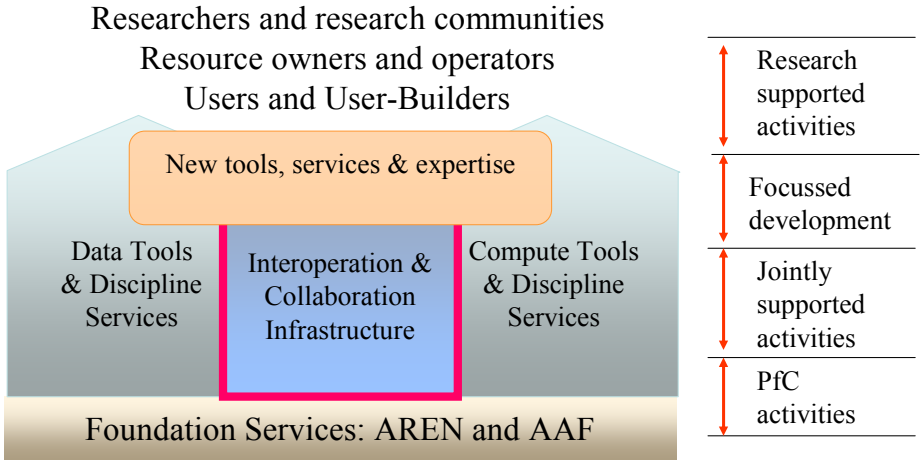
Mutual Dependency

Because the build up over time of national e-Research support and cyber-infrastructure will have many contributors, the PfC investment approach has adopted a "build-with" model.

Therefore a relaxed participation model is proposed where participants do not need to involved in all components, and component members do need to be involved in every activity.

The intention, however, is to build up a balance of activities so that PfC becomes a source of core e-Research services that helps the surrounding set of activities co-ordinate and inter-operate. It is simply the case, that the majority of funding in the e-Research space will always be provided ex-NCRIS.

This ability to attract organisations and other e-Research investment into the overall framework is a key goal, and therefore the level of ex-NCRIS resources contributed to or co-operating within the activities is a key performance indicator.



NCRIS funds national components, and co-funds other components, noting that foundation services are to be mostly member funded

Initial Scope

The overall envelope of notional contribution that PfC is aiming towards is as follows:

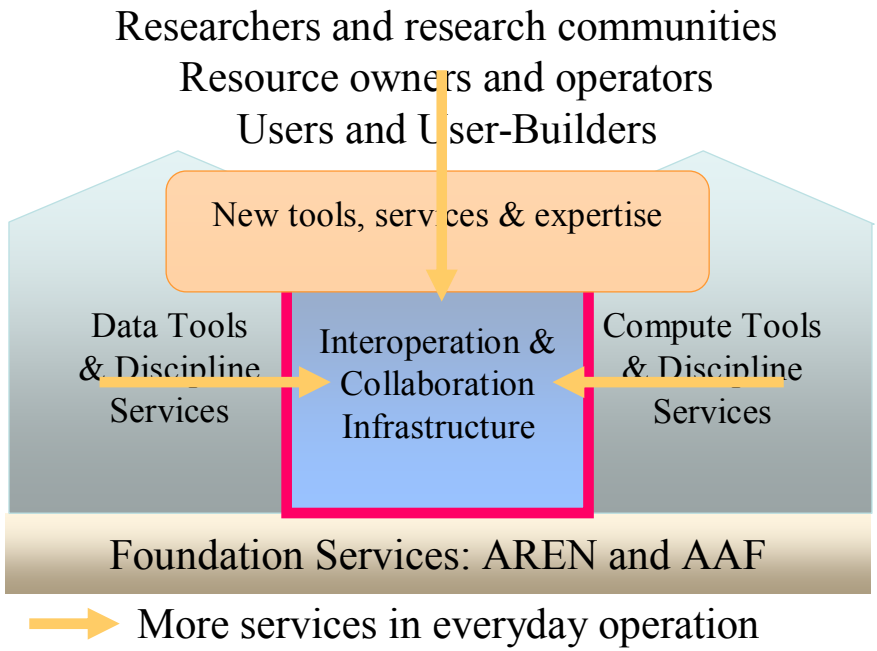
	2007-08 $M	2008-09 $M	2009-10 $M	2010-11 $M	TOTAL $M	Operations $M		Other Cash $M	In-kind EFTs	Total EFTs
Data Management	3.25	5.25	5.75	6.75	21	7		5	110 (27.5pa)	164 (41pa)
Computational Infrastructure	6.5	6.5	6.5	6.5	26	22		16	110 (27.5pa)	132 (33pa)
Interoperation & Collaboration Infrastructure	5	5	5	5	20	14		0	80 (20pa)	208 (52pa)
Foundation Infrastructure	0	2	2	2	6	6		3	N/A	N/A
AeRIC	0.5	0.5	0.5	0.5	2	0		0	N/A	12 (3pa)
TOTAL	15.25	19.25	19.75	20.75	75	49		24	300 (75pa)	516 (129pa)

A primary factor in arranging the investment is that the nature of PfC leads it to prefer to support systems and operational components and to obtain consulting, advisory and expertise components as in-kind contributions.

The arrangement permits organisations to participate without having to fit the mould of a service provider. It also concentrates rather than replicates system and service support expertise and allows for the fact that activities may be best placed for operational reasons near the significant compute or data capabilities they work with.

The intention is that as tools and services develop, they will transition into the inter-operation component to be hosted and supported by suitable service providers. Hence NCRIS funding will need to be provided in a way that allows it (consequently) to also migrate over time.

A second factor concerns the manner in which other e-Research infrastructure investors, such as state governments and research institutions, prefer to channel their investment activities. Most state governments have nominated the existing APAC member as the means for focussing in their region. Some APAC members are in transition to a more generalised e-Research focus and Victoria has created a second entity in addition to the existing APAC member.



However, the expertise required to develop user tools and discipline services and the services identified within the data collections programme do not reside solely within the APAC state partners, and significant resources in associated organisations (such as ANU, CSIRO and the University members of the APAC partners) are not included in APAC's management scope. There is a clear community preference that these additional capabilities for the development of e-Research tools should be harnessed, but that the existing operational service providers are a participant rather than a controller of that process. A second clear preference is that investments relating to tools and data issues be considered independently of the existing compute interests.

This is a major complication, as the benefits that arise from co-location of tools, data and compute resources are high. While network bandwidth is perceived to be able to transport data to compute resources, the future requirement for information services derived from large volumes of data leads inevitably towards the installation of tools and the location of processing capability close to the sources and repositories of data.

The long term vision must be a number of e-Research support facilities that provide for this co-location. However, the urgency for significant investment in data specific services, the immaturity of the tool domain, and the existing contextual gap between data and compute communities, means that this can only be a longer term goal. It is simply too important to begin dealing with research data in its own right, than to ponder eventual improvements, that after all can be accommodated in refurbishment investments, when the requirements are better understood.

The "build-on" strategy does support the achievement of such a vision, and the investment arrangements are intended to broaden and strengthen the role of regional e-Research service providers as the most suitable architecture for reaching this vision within a national cyber-infrastructure.

National e-Research Architecture Taskforce (NEAT)

NEAT will bring together user, platform and discipline groups to work towards the vision:

> *Users are able to work together with familiar tools despite accessing multiple and remote compute, data, analysis, sensor or instrument resources.*

This emergent behaviour is most likely to arise through several 'discipline, problem or task oriented' grids, and the development of common services that can support their simultaneous operation and inter-operation.

This is an ambitious goal if translated across all possible research interests and research disciplines, so that this investment will focus primarily on NCRIS priorities.

However, the underlying motivation is that improved e-Research adoption depends on higher levels of content, application and system inter-operability and the provision of common tools and services that add value to research communities. Without tools and content inter-operability, researchers can only create bespoke components and isolated data sets that require costly case-by-case integration at a later date.

Because of the limited funding, the following priority will apply, as the first is required to extend existing grid infrastructure and the last depends on further development of the NCRIS capability investments:

- first to harden the definition, specification and implementation of the core middleware services required for data and compute system inter-operation;
- second to develop collaboration services that will support the broad e-Researcher community and then harden these for deployment within ICI to provide an initial set of basic collaboration tools;
- third to work with ANDS to identify, implement and harden middleware and tools relevant to data curation and management for new e-Researchers; and
- fourth to work with NCRIS capability and other discipline communities to identify the core middleware services they require, develop operational forms of these services and move them into the national grid.

PfC proposes to fully fund the operational activities within the ICI and ANDS components. Organisations in which many of those services will be located have agreed to contribute additional effort to the developmental, expertise enhancement and user support activities proposed under NEAT.

The rationale for this approach is that e-Research requirements are dominated by the evolving international discipline developments in middleware and tools, and only those involved in such activities can offer advice on the directions likely to arise.

A gap exists and will continue to exist, where service providers are focussed around system and middleware, inherently seeking generic services, and researchers are focussed into specific solutions and specific tools by international developments.

The goal of NEAT is to close that gap by bringing together service providers, researchers in infrastructure services, and researchers in application spaces to guide the continuous evolution and development of the operational cyber-infrastructure.

The rationale from regional service providers and research groups, is that separate piecemeal approaches to every NCRIS capability (as an example) is most likely to alienate everybody involved and harm their own goals.

Hence, while the regional service providers are funded to undertake such work and middleware oriented research groups are enhanced by engagement with applications, NCRIS can achieve significant overall benefits by developing a national orientation that engages these groups in coordinated (and preferably concerted) activities.

The particular importance to NCRIS is that the process is structured to identify, trial and test the middleware and user tools that can be adopted as part of the developing national cyber-infrastructure. The intention will be to migrate such tools into operational services.

Three kinds of activities are contemplated:

- Projects that 'adopt, adapt and influence' selected middleware components into demonstrable user and discipline focussed tools and services.
- Investments in nationally significant expertise that adds value to infrastructure development and aligns with strategic needs identified by AeRIC.
- Resources for service providers to sustain additional tools and services as they are deployed into the operational infrastructure.

The funding in these will vary over time, as will the relative emphasis between general user tools and more generic discipline services. Resource allocation will be based on the opportunities that are most likely to create new services or to enhance existing services within the operational infrastructure.

Trial Services

This is the primary means by which new tools and services will be identified, trialled and developed for deployment as an operational service.

Because of the steep learning curves in middleware, the number of technologies that need to be integrated to develop a service, and the need to develop understanding between communities, projects are expected to be generally of two years duration.

Projects will only be considered in response to the infrastructure development timeline published by NEAT.

Projects that deploy resources within PfC components which do not require additional funding can be approved by the relevant component director on the recommendation of NEAT.

Projects which commit further funds, will need endorsement by NEAT for approval by AeRIC.

Expertise Development

Activities in this category will build nationally recognised centres of support for key technologies identified by NEAT.

Such expertise will be expected to assist the adoption, adaption and influence of global standards, components and middleware to create discipline oriented formats, tools and services. The activities will be expected to participate with global peers in developing standards, components and middleware where appropriate.

Some of these centres are likely to be associated with the organisations involved in other parts of the investment plan; however, some may be associated with e-Research 'user' communities. Proposals for expertise support will be approved by AeRIC.

Tools Environment

As new infrastructure components and particularly as an increasing variety of user tools are made operational, funding for additional expertise and operational cost needs to shift to the operators of the infrastructure.

Recommendations in this category will be negotiated with relevant services providers and would result in a change to the budget in that component to support the newly developed services.

Budget

The budget of NEAT activities is $10M from NCRIS combined with up to twice that in participant resources.

The expectation on leverage is justified from the full funding of operating costs of core services by NCRIS and for the fact that participants in NEAT activities will be self selecting by virtue of their existing activity and interest levels in those activities.

As an indicator of the challenge facing the NEAT process, the following immediate opportunities for NEAT have been identified during the facilitation process:

- expertise build up to support campus network harmonisation with AREN;
- development of AAF integrated collaboration tools;
- trial deployment of the iRODs successor to SRB;
- hardening and extension of data transfer services;
- hardening and extension of job submission fabric;
- installation and operation of a re-configurable high capacity research network;
- development and hardening of a simple meta-data capture tool;
- development of initial forms of ANDS technical services;
- transfer into operational mode of ARCHER tools;
- development of initial workflow services;
- establishment of a national Nimrod service;
- development of reference architectures for research data management solutions;
- development of national reference models for intellectual property (IP) management in research data;
- development of national reference models on roles and rights for the AAF;
- any of the AREN developments contemplated but not funded by PfC; and
- the integration of AAF protocols into any number of applications.

While this appears initially as a daunting list, the developments are not on-going, and not limited to Australia (ie some of these are global issues) and they also have some temporal relationship to the overall development of the e-Research infrastructure.

The maximisation of non-NCRIS funding in these activities is vital if flexibility and a satisfactory overall rate of progress is to be achieved.

Contractual Overview

The investment plan defines a variety of activities. DEST will enter into the requisite contractual arrangements, as follows:

Service Cluster		Form	Contributions	Management	Duration	Resource and IP[5] Ownership	Service basis
National e-Research Architecture Taskforce (NEAT)		For each activity: Deed of grant, Project plan	AeRIC: scoping and review Others: resources, services and service developments	NEAT, Project leader	Multiple 2 year activities	Participants; IP to participants	Strategic allocation on an architecture development feature against a discipline or NCRIS capability need
Australian National Data Service (ANDS)		NCRIS agreement, Business plan, To be determined	AeRIC: scoping and review Others: resources, services and service developments	ANDS Director, Governing Committee, Annual Plan	4 years ongoing	Partners; IP to public domain	Public access, merit assignment on a research community basis, or on a full cost basis
National Computational Infrastructure (NCI)		NCRIS agreement Business plan, NCI Project Partner Joint Venture (JV)	AeRIC: scoping and review third parties: funding Hosts: resources, services and service developments	NCI Project or Host Directors, Steering Committee, Annual Plan	3-4 years ongoing or fixed	ANU, RNSPs[6], or third parties; IP to public domain	Strategic allocation by capability, merit assignment to individual researcher, or on a full cost basis
Inter-operation and Collaboration Infrastructure (ICI)		NCRIS agreement Business plan, ICI Partner JV ICI Affiliate contract	AeRIC: scoping and review RNSPs: resources and services third parties: resources to implement and deliver services	ICI Director, Governing Committee, Annual Plan	4 years ongoing	RNSPs or third parties; IP to public domain	Available free to all researchers with access to a connected research resource
Foundation	AREN	Deed of grant, Annual reports	DEST: scoping and review AARNet: extended services	Annual review	Ongoing	AARNet; IP to AARNet	Funded by subscription or fee for use
	AAF	Deed of grant, Business plan	AeRIC: scoping and review AAF: services, and service and market development	Annual review	Ongoing	AAF; IP to public domain	Public access, available to all researchers, funded by subscription, or fee for use

[5] IP refers to Intellectual Property developed under the NCRIS investment

[6] Regional and National Service Providers

PART SEVEN Background Context

Relationship to UK e-Science Directions and Plans

The following quotation is the entire executive summary to "Developing the UK's e-Infrastructure for Science and Innovation", a Report of the UK Office of Science and Innovation (OSI) e-Infrastructure Working Group[7].

"The growth of the UK's knowledge-based economy depends significantly upon the continued support of the research community and in particular its activities to engage with industry and to apply its world-leading innovations to commercial use. A national e-infrastructure for research provides a vital foundation for the UK's science base, supporting not only rapidly advancing technological developments, but also the increasing possibilities for knowledge transfer and the creation of wealth.

"With e-Infrastructure requirements common across research disciplines, across Government departments and across sectors, such an e-infrastructure can further these and other key national objectives and indeed, through its support of world-leading research, help answer some of the 'grand challenges' facing the UK and the wider world, such as climate change, an ageing population and the combating of disease.

"Technology is in large part driving globalisation – through increased specialisation, greater market integration and the removal of barriers, the sharing of knowledge and expertise and through its impact on worldwide economic growth. There is the danger, however, that in the new global market in which investment in technology and reward for innovation will increasingly determine a country's economic performance, the developing countries will 'leap frog' technological progress and overtake the world's more established economies.

"Evidence already suggests that with massive investments being made in infrastructure development by rapidly industrialising countries such as China, India and South Korea, and by other more developed competitors, the UK is beginning to lag behind these worldwide advances, causing damage to its international competitiveness as well as its global leadership in research.

"In March 2000, the EU Heads of States and Governments agreed to make the EU 'the most competitive and dynamic knowledge-driven economy by 2010'. While progress has been made in achieving the 'Lisbon goals', there is growing concern that the reform process is not going fast enough and that the ambitious targets will not be reached. Greater investment in the national e-infrastructure and the more coordinated approach to its development, detailed in this report, would therefore support not only the UK's vital national objectives but also its international commitments.

"In the face of these worldwide developments, the UK needs a national e-Infrastructure capable of meeting the needs of UK research and researchers in the digital age and the needs of the UK in a global market.

"Elements of a UK e-Infrastructure have grown over the years and indeed have helped secure the current standing of UK research, supporting vital developments in the pharmaceuticals, defence, information and media, financial services and other industries. However, with rapid advances in technology and the increasing dependence of the research community on data-intensive forms of research, such a position is not sustainable without a step-change in national provision and concerted action towards e-infrastructure development.

"A national e-Infrastructure needs: the means of producing, managing and preserving vast amounts of digital data; sophisticated means of accessing an ever-increasing range of electronic resources of all kinds; technologies and structures to support dynamic and virtual communities of researchers; unprecedented network, grid and computational capacity; and the necessary national services and systems to ensure safe and secure access to resources. We believe that these and other requirements presuppose not only a high level of integration and coordination, but also, in key areas, intervention at the policy level.

"Underpinning all these requirements is the need for the adoption of agreed technical and other standards, for the appropriate training and skills to ensure that UK researchers from all disciplines can compete in a worldwide market, and for strong

[7] The OSI e-Infrastructure Working Group was formed to explore the current provision of the UK's e-infrastructure and to help define its future development. Formed in response to the 'Science and Innovation Investment Framework 2004 – 2014', which was published by the Treasury, the DTI and the DfES in 2004, it is made up of senior representatives from the Research Councils, JISC (Joint Information Systems Committee), RIN (Research Information Network) and the British Library.

coordination between government, funders, research and development agencies, service providers, and universities and research centres.

"The UK has a world-class research base, second only to the USA in global excellence as measured by citations. We believe that a national e-infrastructure built on the foundations outlined above and detailed in this report will be one that will enhance the global standing of UK research, maximise the immense potential of new technologies for knowledge transfer and help the UK achieve its wider social and economic goals."

The UK vision for a national e-Infrastructure is described as follows:

"The UK's e-Infrastructure should provide researchers with:

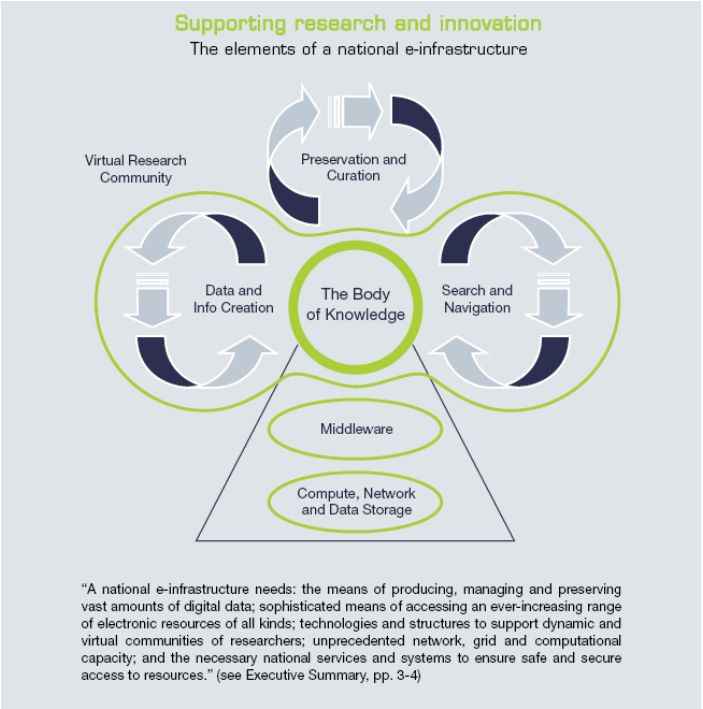
- Access to the systems, services, networks and resources that they need at the point that they need them;
- Facilities to discover resources easily and use them appropriately;
- Confidence in the integrity, authenticity and quality of the services and resources they use;
- Assurance that their outputs will be accessible now and in the future
- A location-independent physical infrastructure for combining comp;utation and information from multiple data sources;
- Advanced technologies to support collaborative research; and
- The training and skills needed to exploit the services and resources available to them.

"The e-Infrastructure should allow researchers to:

- Exploit the power of advanced information technologies and applications to continuously enhance the process of research itself;
- Collaborate and communicate securely with others, across disciplines, institutions and sectors;
- Maximise the potential of advanced technologies to support innovation and experimentation;
- Share their research outputs with others and re-use them in the future; and
- Engage with industry in support of wider economic goals.

"The e-Infrastructure must enable:

- The growth of knowledge transfer and the development of the commercial applications of research outputs;
- Research funders to track the outputs from the research they fund;
- The protection of individuals' privacy and work, within regulatory, legal and ethical constraints;
- The protection of intellectual property and rights management;
- The preservation of digital information output as a vital part of the nation's cultural and intellectual heritage".



In relation to this PfC Investment Plan, it is obvious that the U.K. view is highly advanced in its focus around the use and re-use of content. The PfC Investment Plan is more focussed on building the infrastructure, centres of expertise and day-to-day operational services that can provide a foundation for a more data and content inter-operation view. We need to aim higher.

Alignment with Cyber-Infrastructure Development by the US National Science Foundation

The latest draft report from the Office of Cyberinfrastructure in the US National Science Foundation (NSF) provides the recommendations quoted below.

Each goal is followed by an indication of how the PfC investment relates to that goal.

The numbering is from the NSF cyber-infrastructure report.

(a)	Provide communities addressing the most computationally challenging problems with access to a world-class high performance computing (HPC) environment.
PfC	Continue the investment in the National Facility, acknowledge it is a specialist service and that its mission is to operate with an expert user orientation (HPC).

(b)	Broaden access to state-of-the-art computing resources, focusing especially on institutions with less capability and communities where computational science is an emerging activity.
PfC	Develop an integrated co-investment plan for regional compute and data support services (HPC, Data), and support a wider range of new e-Research users through a separate and strong focus on commodity tools and services (Tools).

(c)	Support the development and maintenance of robust systems software, programming tools, and applications needed to close the growing gap between peak performance and sustained performance on actual research codes, and to make the use of HPC systems, as well as novel architectures, easier and more accessible.
PfC	As a research activity this is outside of NCRIS scope; however, NCRIS will require national pools of expertise and these can be tasked to provide support (where agreed) to research endeavours likely to lead to advanced application demonstrators (Support).

(d)	Support the continued development, expansion, hardening and maintenance of end-to-end software systems – user interfaces, workflow engines, science and engineering applications, data management, analysis and visualization tools, collaborative tools, and other software integrated into complete science and engineering systems via middleware – to bring the full power of a national cyber-infrastructure to communities of scientists and engineers.
PfC	Strengthen and focus the community support activities of the current APAC Grid programme into user/discipline services activity using NCRIS capabilities as a prime motivator.

(e)	Support the development of the computing professionals, interdisciplinary teams, enabling policies and procedures, and new organizational structures such as virtual organizations, needed to achieve the scientific breakthroughs made possible by advanced cyber-infrastructure, paying particular attention to the opportunities to broaden the participation of underrepresented groups.
PfC	Develop the Australian Access Federation.

(f)	Support state-of-the-art innovation in data management and distribution systems, including digital libraries and educational environments that are expected to contribute to many of the scientific breakthroughs of the 21st century.
PfC	Initiate an Australian National Data Service to support these objectives.

(g)	Support the design and development of the cyber-infrastructure needed to realize the full scientific potential of NSF's investments in tools and large facilities, from observatories and accelerators to sensor networks and remote observing systems.
PfC	Pursue campus infrastructure network harmonisation, accelerate extension of a national grid systems/middleware effort to additional parties.

(h)	Support the development and maintenance of the increasingly sophisticated applications needed to achieve the scientific goals of research and education communities.
PfC	Initiate a programme focussed in the tools and discipline support area.

(i)	Invest in the high-risk/high-gain basic research in computer science, computing and storage devices, mathematical algorithms and the human/cyber-infrastructure interfaces that are critical to powering the future exponential growth in all aspects of computing, from hardware speed, storage, connectivity and scientific productivity.
PfC	In general, out of scope; however, support can be provided to configure and operate infrastructure as part of research involving advanced demonstrations.

(j)	Provide a framework that will sustain reliable, stable resources and services while enabling the integration of new technologies and research developments with a minimum of disruption to users.
PfC	Resource a strategic oversight for the PfC and other NCRIS informatics activities.

Alignment with the Report of the e-Research Coordinating Committee

The executive summary of the report of the e-Research Coordinating Committee (eRCC) made the following key points, presented as excerpts from the text.

The e-Research Strategic Framework

The key issues addressed in the report are:

- the need for leadership to drive cultural change;
- fostering engagement, participation and cooperation;
- research, development and deployment of ICT solutions for e-Research;
- skills acquisition;
- support for researchers through a national e-Research Centre;
- access, authentication and authorisation;
- support for the physical resources, software and services that enable e-Research activities (otherwise known as the e-Research 'fabric');
- data management and accessible databases; and
- coordination.

A Vision for e-Research in Australia

The report identifies the following vision for e-Research:

"Australian researchers will enhance their contribution to world-class research endeavours and outcomes, through the use of advanced and innovative information and communications technologies.

"The vision encourages researchers to participate in the transformation process being enabled by ICT, as it offers the power to undertake research on a scope previously unattainable, to work collaboratively and globally in a way not previously possible, and to improve existing research."

The need for leadership and coordination

"Progressing the e-Research agenda is as much about people as it is about technology, because e-Research is challenging existing research practices and cultures. There is a need for strong leadership to articulate the vision and engage key players in taking the strategy forward in a coordinated manner.

"Overseas experience, particularly in the UK, strongly suggests that success in engaging the research community in e-Research is associated with the appointment of a 'champion', a well-respected member of the research community, recognised by government, the research sector and business as the visible face of e-Research."

PfC: AeRIC would provide a suitable forum within which one or more such 'envoys' might be appointed, supporting an association with broad discipline interests.

Building on the Current Research Base

"The take-up of e-Research will be reliant on the availability of fully functional, scaleable applications, user interfaces and data management and administrative services, defined collectively as 'operational services'.

"To date, the Australian Government has invested in a range of innovative projects under the auspices of the Australian Research Information Infrastructure Committee (ARIIC), the Advanced Networks Programme (ANP), the Australian Research Council (ARC), the National Health and Medical Research Council (NH&MRC), the CSIRO and other organisations. While the outcomes of these projects are informing the development of operational services, further research and development to support Australian e-Research capabilities is essential. In particular, Australian participation in international standards-setting should be encouraged, and the Committee recommends that funds be dedicated to supporting these activities."

PfC: The Interoperation and Collaboration Infrastructure provides the cohesive support of fully operational services and the Application and Discipline Services is a programme which directly addresses this requirement.

Skills Acquisition

"Many of the technologies that support e-Research are at the cutting edge of developments world wide. Although Australia does have world-class leaders in some of these technologies, there is a nation-wide lack of skills to support the rapid take-up of e-Research, and. recommendations focus on developing:

- practical skills for researchers who want to use e-Research applications;
- skilled professionals who can operate across research domains to help researchers use e-Research applications; and
- experts in ICT and information management who are capable of undertaking research and developing new e-Research platforms and applications.

"There is a need for education and research opportunities to equip highly motivated students to choose careers in the technologies which support e-Research. A number of Australian universities have already introduced course modules that focus on aspects of e-Research into their undergraduate degree programmes."

PfC: This is mostly unaddressed, other than through incidental development arising through project work.

Skills Transfer and Support for Researchers through an e-Research Centre

"In the UK, the success of the government funded e-Science initiative was due in large part to the establishment of several e-Science centres which provided a focal point for research, development, support and outreach into e-Science.

"Similarly, the Committee recommends the establishment of a national e-Research Centre comprised of several nodes located in regions of research intensity. Their primary role would be to act as change agents locally and nationally to embed e-Research methodologies in research practices. The nodes would provide outreach programmes and support services for researchers and serve a critical role in skills development, skills transfer and the development and deployment of e-Research applications."

PfC: The investments will create a new centre of expertise in data management skills, enhance grid operational skills, sustain HPC skills and develop over time a broadly based skill set related to e-Research solution development.

Access for Researchers

"A key enabler for e-Research in Australia is to give researchers seamless access to resources, including each other. Resources are distributed in various areas of Australia and overseas and include:

- digital data repositories;
- scientific facilities, instruments and sensors;
- computational facilities; and
- high-speed telecommunications networks.

"The Committee has identified the need to put in place a national, robust identity provision solution that will allow researchers' identity to be authenticated, and their access authorised, to distributed resources via a single sign-on system."

PfC: The existing APAC grid will be challenged to develop a more sophisticated standardisation of its processes and services, and to outreach to similarly connect resources beyond the its participants.

The AAF will be established to provide the required identity solution.

Support for the Research Fabric

"The 'e-Research fabric' refers to the physical resources, software and services that link, or are linked to enable, e-Research activities. They include broadband networks, middleware services, computing capability, scientific instruments, and digital data repositories. Linking the physical resources can significantly add capacity to the e-Research infrastructure."

"…there will be an ongoing need to extend broadband capability to more remote institutions in the network and to other key sources of data and resources, including research agencies, cultural institutions and strategically important international partners.

"The availability of high performance computing capability is critical to support a number of e-Research applications. As a result of substantial Government investments in recent years, the Australian Partnership for Advanced Computing (APAC), a number of universities and publicly funded research agencies provide high performance computing facilities in various parts of Australia. Many of these computing facilities are networked and therefore can provide distributed computing capability. In terms of implementing the e-Research agenda, it is important that as

many such facilities as possible be available to service researchers' increasingly advanced computing needs."

"…The full utilisation of advanced ICT infrastructure is increasingly dependent on computer software, termed middleware, that links the ICT resources that users need. Middleware provides the common set of services and tools that allows researchers and applications to treat the data repositories, computing, and other disparate resources as if they were one large virtual facility.

"While there has been considerable progress in middleware development and deployment both within Australia and overseas, the Committee recommends that arrangements should be put in place to reduce duplication of effort and deliver widespread efficiencies."

PfC: The operational grid will be challenged to extended to all significant compute, data, instrument and sensor resources as are nationally significant.

Because funding is limited, middleware development will be limited to components required by application services in NCRIS capabilities and other nationally important disciplines and specifically related to components expected to migrate into operational services in the short term.

Managing Data

"Managing data effectively is essential to support the full cycle of research endeavour, from research concept formulation and scoping to the research activity itself, to dissemination of the results of research.

"The Committee identified the data management needs of researchers as being:

- data collection and generation;
- data storage and the physical management of stored data;
- evolution of standards and protocols to facilitate the storage, use and interpretation of data;
- access to data; and
- long term archival and preservation of data and policy for retaining and discarding of data.

"Research domains globally are generating unprecedented quantities of data and the issue of data management is increasingly a critical one. There is need to balance the likely accessibility needs of researchers to data against the likely costs to the research sector and the wider community of long term data retention.

"The Committee recommends that the Government convenes a working group to develop an Australian Research Data Strategy."

PfC: AeRIC provides a forum to develop the Australian Data Strategy.

Because funding is limited, the primary focus of the investments in data management is towards supporting researchers in data collection and generation, and developing services that assist access to data. Some support is provided to storage and retention for critical collections (such as irrecoverable time-based studies).

The problem of the explosion in data retention is not directly addressed; however, a co-operative framework of regional data retention services will be established.

Coordination, Oversight and Resources

The report recommends that "an e-Research Committee be established to replace the existing Coordinating Committee in order to implement the e-Research Strategic Framework over a period of five years. The new Committee would provide the strategic direction, drive and engagement and coordination of effort of research groups involved in e-Research. It would also take account of existing Systemic Infrastructure Initiative (SII) advisory structures and develop a close working relationship with the NCRIS Committee and its facilitators."

PfC: AeRIC would satisfy this requirement.

The requirement strengthens the case of a government operated lead activity as opposed to an independent incorporated entity, which by definition must have less attachment to government policy and direction setting.

As the eRCC proposal was largely concerned with expertise development and as PfC will spend at most $20M pa including hard infrastructure and service implementation, most of the needs can only be assessed as partly met and some must be considered unmet. There is at least a $15M pa shortfall in the areas identified for support by the eRCC compared to the contribution PfC will make to those areas.

Comments on Transitions in Service Level Appropriate to Australian Infrastructure Services

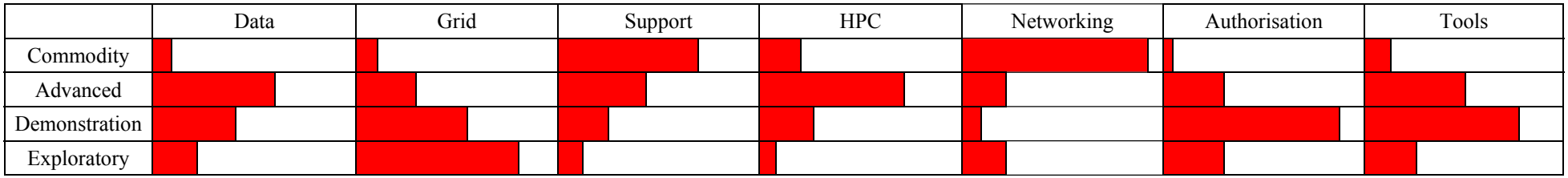
e-Research will need to become an increasingly normal part of research activity if the vision is to be realised, and PfC will need to invest in ways that develop the appropriate emergent capabilities from the results of many investments by many participants.

Hence the investment plan sets out to act on the difference between the current state and the proposed future state for each of these emergent capabilities. It does that by ensuring that necessary and appropriate e-Research services exist and are universally accessible.

Because broad adoption always depends on commoditisation, the plan naturally considers where and how 'commoditised'[8] e-Research services can be developed. It is important to recall that leading edge e-Research cannot be commoditised by definition. However, it is also the case that it should build on commodity services where ever possible. So a better understanding around commodity services needs to be developed.

To do that, it is useful to consider services in several classes: Commodity/generic, Advanced/specialised, Demonstration/prototype, Exploratory/research. The services in these classes could be expected to require increasing levels of user expertise for their effective deployment, have decreasing quality of service levels and perhaps decreasing availability, and involve different partners and arrangements for co-investment.

Considering the areas above, and limiting a view to e-Research services, the overarching investment goals can be established by examining a representative pattern of current activity; which approximately might be as follows.



Broad e-Research adoption depends on some of these patterns being changed.

Data	At present, only some enterprise research data and some community reference collections have a permanent home. Many disciplines are creating purpose built data retention and access services, operated on a co-operative basis by research groups. Most leading edge analysis is bespoke and user based, though the provision of third party analysis tools and methodologies is increasing (and prevalent where applicable on the desktop in disciplines such as engineering).
Delta	Simple retention and curation tools need to be broadly available, standard practice needs to be promulgated and understood and help made readily available. The development and support of relevant extraction, reduction and analysis functions as part of each data service is an essential step required in the transition to a future of information services rather than data services.
Implications	Commodity hosting of these services is needed, for data already managed and more importantly for data yet to be managed, further demonstration activities will also be required

[8] Where commoditised in this context means: able to be used with generally available discipline knowledge and without expert or deep service specific knowledge or sophisticated technological understanding.

Grid	At present, nearly all grid services are provided in advanced or demonstration modes and users are regularly exposed to the details of the technologies. Some leading discipline specific examples are in daily operation, but not yet in what could be considered a commodity user form.
Delta	Unfortunately, little change in this state is expected over the next 5 years as the technology is under evolutionary development. This is a major constraint on the investment plan.
Implications	The majority of investment in the grid area will need to target expert users, advanced research communities and can only realistically support the "builders" of user facing services in the short term.

Support	It is important to note that the majority of staff providing ICT support to researchers operate within well managed ICT support services, but that the technologies that PfC might deploy, are not currently within the scope of those services. Consequently, much of the support activity relevant to PfC operates in services aimed at expert users, or as an adjunct to research teams and communities, and is either outside of the rigour of commodity service provision or difficult to access as an in-expert user. Overall, there is a significant shortfall in relevant and available expertise.
Delta	Centres of expertise have been built in areas such as HPC and networking, and to an extent in data (where investments have been made). Stronger investment is required in relevant expert groups and strategies are needed that can migrate that expertise into providers focussed on broad support and service delivery.
Implications	National co-operative provisioning and use of platform technology expertise is crucial. Support functions will be best placed with their related services and providers, rather than in an independent activity, in order to better cope with the steep learning curves required. Support will only be able to be treated more systematically in areas where stability can be achieved in the technology platforms and the tools adopted by the various disciplines to support their e-Research activities.

HPC	A peak capability exists, shoulder capabilities are growing, and many independent other resources exist. System wide brokerage of resource use and work migration and co-ordination remain difficult.
Delta	Peak computing services will remain advanced services for the foreseeable future and are not amenable to commoditisation. The Peak facility will always need to manage itself towards a limited market. The overall investment in HPC needs to better meet the needs of a broader user base by provisioning compute support for tools and analysis services. PfC needs a coherent compute fabric so that successful tools and services can migrate or be replicated as part of supporting broader e-Research adoption.
Implications	The peak capability is well understood from a mission perspective, along with its related processes and investment requirements, and should be protected from demands for generic or commodity services. An investment outside of peak capability is needed to meet the need for a generic computing environment for commodity tools and analysis services.

Networking	Network technology is very well developed, so that 'commodity' services with high levels of service are the norm, and indeed provided. In the case of the AREN, many non-academic researchers are not connected, tolls exist in institutional frameworks, and high bandwidth transfers remain difficult. Overall research traffic volumes do not yet justify the network investment, a problem which needs to be addressed by encouraging research data interchange.
Delta	Effective access must be provided to researchers outside of the AARNet members. Institutional network infrastructures and policies need to be harmonised with the vision for the AREN. Some increase in advanced and demonstration quality activities is needed to develop the future research support focus of the network.
Implications	An expert group should be established to assist the harmonisation of campus infrastructures, an investment in demonstrators that lead to higher use is crucial, an outreach network (with possibly different QoS) should be established to open AREN to all appropriate Australian researchers (for on-net traffic).

Authorisation	Inter-institutional (and hence regional, national and international) authorisation services with reasonable usability characteristics, are only just now being constructed anywhere in the world.
Delta	A shift to the provision of a set of simple core services in a commodity service mode is crucial, further demonstration activities will also be required as this area is expected to evolve rapidly. Modification of applications and e-Research support tools will be needed to interface with the authorisation services, if uptake is to occur.
Implications	The foundation of the AAF as a core e-Research service provider is essential for broader e-Research adoption. On-going investments will be needed to assist that process, which should focus on developing a centre of expertise around the service, providing exemplar collaboration tools and services, and an outreach activity aimed at adapting discipline specific e-Research tools to interface with the service.

Tools	At present, analysis activities often require a researcher to understand the tools they use, the tools others use, the data sources, the specific compute engines as well as many of the intervening systems. Many disciplines have developed some common methods and tools, often on a best effort basis. Little co-adoption across disciplines is evident, with a notable exception in the spatial data area. Standards-based web and messaging solutions appear to meet some of the research collaboration needs, and can be expected to provide a variety of highly commoditised collaboration tools.
Delta	The fundamental resource in this area is human, and its availability is limited. Hence some choices in middleware and tool components need to be made if more commodity like services are to be supplied to a larger set of e-Researchers. This should be through an explicit investment in building expertise, support and the provision of capacity around nominated components. PfC should focus on functionality and tools for the builders of cyber-infrastructure and work with other NCRIS investments to identify and support functionality and tools for users. A nationally coherent computing/data management fabric that could host application selected tools and components would significantly increase their adoption.
Implications	Investment is needed to establish a national compute/data fabric, which would need to include existing regional providers and major research institutions. Strategies for effectively deploying and rapidly developing expertise levels right across e-Research interests is a crucial requirement of the investment plan.

List of Acronyms

AAF	Australian Access Federation
ANDS	Australian National Data Service
ANG	Australian National Grid
ANP	Advanced Networks Programme
ARCHER	Australian ResearCH Enabling enviRonment
AREN	Australian Research and Education Network
CRC	Cooperative Research Centres
CT&T	Computational Tools and Techniques
EFT	Equivalent Full Time
eMII	e-Marine Information Infrastructure
eRCC	e-Research Coordinating Committee
EU	European Union
HPC	High Performance Computing
ICI	Interoperation and Collaboration Infrastructure
ICT	Information and Communications Technology
IP	Internet Protocol (also IPv6 IP version 6)
IT	Information Technology
JV	Joint Venture
LHC	Large Hadron Collider
LIEF	Linkage Infrastructure, Equipment and Facilities
MAMS	Meta Access Management System
MNRF	Major National Research Facility
NANO	Nanostructural Analysis Network Organisation
NatFac	National Facility
NCRIS	National Collaborative Research Infrastructure Strategy
NDN	National Data Network
NEAT	National e-Research Architecture Taskforce
AeRIC	Australian e-Research Infrastructure Council
PfC	Platforms for Collaboration
PKI	Public Key Infrastructure
PMSEIC	Prime Minister's Science, Engineering and Innovation Council
QoS	Quality of Service
R&E	Research and Education
RIN	Research Information Network
RNSP	Regional or National Service Provider
SII	Systemic Infrastructure Initiative
TOR	Terms of Reference
VeRSI	Victorian e-Research Strategic Initiative
VO	Virtual Organisation

Organisations

AARNet	Australian Academic Research Network (also AARNet3)
ABS	Australian Bureau of Statistics
ac3	Australian Centre for Advanced Computing and Communications
ACcESS	Australian Computational Earth Systems Simulator
ANSTO	Australian Nuclear Science and Technology Organisation
ANU	Australian National University
APAC	Australian Partnership for Advanced Computing
ARC	Australian Research Council
ATNF	Australian Telescope National Facility
AVCC	Australian Vice-Chancellors' Committee
BoM	Bureau of Meteorology
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DCITA	Department of Communications, Information Technology and the Arts
DEST	Department of Education, Science and Training
eRSA	e-Research South Australia
GA	Geoscience Australia
IMB	Institute for Molecular Bioscience, The University of Queensland
iVEC	Interactive Virtual Environment Centre
JISC	Joint Information Systems Committee (UK)
NAA	National Archives of Australia
NHMRC	National Health and Medical Research Council
NLA	National Library of Australia
NSF	National Science Foundation (USA)
QCIF	Queensland Cyber Infrastructure Foundation
SABRENet	South Australian Broadband Research and Education Network
SAPAC	South Australian Partnership for Advanced Computing
TPAC	Tasmanian Partnership for Advanced Computing
VERNet	Victorian Education and Research Network
VPAC	Victorian Partnership for Advanced Computing

Software

Blast	The Basic Local Alignment Search Tool (used in bioinformatics)
FTP	File Transfer Protocol
Glite	Lightweight middleware for Grid computing
Globus	A very commonly deployed US middleware suite
GridFTP	Protocol extensions to FPT for the grid
iROD	i Rule Oriented Data Systems (successor to SRB)
NimRod	Tools for distributed Parametric Modelling
SRB	Storage Resource Broker
SRM	Storage Resource Management
Unicore	HPC oriented grid middleware suite