

Platforms for Collaboration

Dr Rhys Francis

NCRIS Facilitator (Platforms for Collaboration)
CSIRO, Level 11, 700 Collins Street, Docklands VIC 3008

Introduction

E-Research involves the application of advanced and innovative information and communication technologies to any research, as well as the new research techniques and new fields of research made possible by those technologies.

The spreading use of e-Research methodologies, including data mining and simulation, means that e-research infrastructure includes hardware relating to computing, data storage and communications networks; and also includes complex models developed with considerable investments to examine large research challenges, such as climate change, water management and land use; and the corpus of data collected across experiments and studies.

The Platforms for Collaboration (PfC) capability of the National Collaborative Research Infrastructure Strategy (NCRIS) 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 e-Research Coordinating Committee:

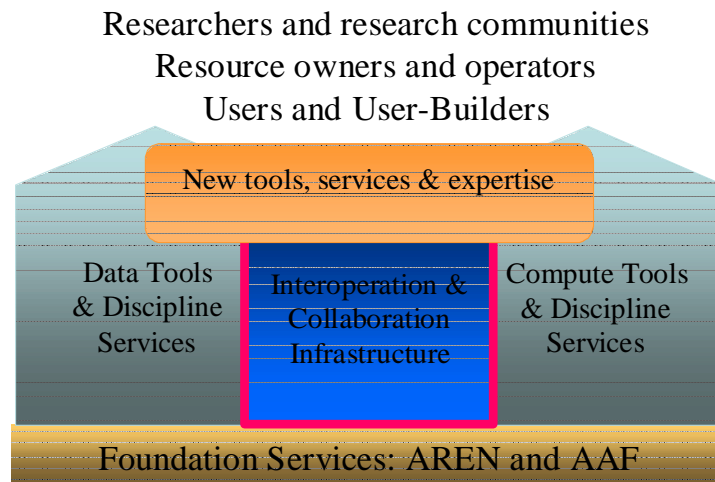
"Australian researchers will enhance their contribution to world-class research endeavours and outcomes, through the use of advanced and innovative information and communication technologies."

This paper provides a brief introduction to the infrastructure and services to be developed within the NCRIS framework.

Previous investment into e-Research has included:

- 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)

Following extensive consultation, the recommendations accepted by NCRIS relating to Platforms for Collaboration involves six major activities of which four require the establishment of organisations and three require NCRIS contractual arrangements for implementation.



These are depicted in the diagram above, and involve:

- Continuation of the AREN
- Development of the Australian Access Federation
- Development of a National Data Management Infrastructure
- Extensions to the National Computational Infrastructure
- Development of a National Interoperation and Collaboration Infrastructure
- Establishment of a National e-Research Architecture Taskforce

In addition, a national e-research council will be established to provide strategic advice to government and guide the coherent development of the infrastructure.

The Australian e-Research Infrastructure Council

To oversee the e-Research landscape, an Australian e-Research Infrastructure Council (AeRIC) is to be established which 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
- recommend the manner in which funds will be provided to e-Research activities in accordance with the investment plan, and subject to annual re-assessment with NCRIS and DEST
- review progress and the contribution of activities within the e-Research funded by NCRIS, and recommend variations to funding agreements to NCRIS 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:

<ul style="list-style-type: none">▪ data capture, management, retention, publication, discovery and re-use▪ data analysis and visualisation▪ computational modelling	<ul style="list-style-type: none">▪ collaboration systems▪ grid inter-connectivity▪ trust and access management▪ networking
--	--

The Chair of AeRIC will have a broad view of national e-Research requirements and the members (about ten members are envisaged) will be drawn from the stakeholders and participants in e-Research infrastructure; including providers and users of e-Research services and capabilities.

An Executive Director will be appointed to identify and investigate integrative services of value across NCRIS capabilities and other research activities and to develop an annual business plan for platforms for collaboration that extends and improves national e-Research infrastructure over time.

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.

Australian Research and Education Network

Previous DEST investment in the AREN already provides research institutions with early access to world class high bandwidth research connectivity. Because 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.

However, research traffic volumes need to rise further to demonstrate that such access is provided at a below market price by virtue of the investment in that infrastructure. Also, the bedding down of the existing infrastructure in terms of its service offerings, and their policy and price integration with campus infrastructures and service policies needs to occur. Many research benefits could be achieved by a better alignment between campus and regional infrastructures and policies and the AREN.

As communication is a normal cost of any business or activity, the investment plan has taken the position that any direct subsidisation of the AREN's operational activities would be a form of cost shifting and therefore has not been proposed.

Various options do exist for network extension and technology refreshment for an AARNet4 will be required at some stage. Nevertheless, the plan has focused on connectivity investments into the higher level services that provide value over the top of the AREN rather than in investments into the AREN backbone itself. The basis for this has been the view that connecting researchers and resources and increasing traffic levels is needed to establish a clear business case for further core investment.

Australian Access Federation

In authorisation systems, the lack of harmony at a policy level and the lack of interoperability at a technical level across research institutions, is a major impediment to collaboration. While manageable by work-arounds in human timescales, the situation must be remedied if electronically mediated collaboration is to become routine.

DEST have provided funding through the Strategic Infrastructure Initiative to demonstrate and trial solutions and have provided a further \$4.8M in 2007-09 to start-up a trust federation, the best means currently available for tackling the problem. The trust federation will be called the Australian Access Federation (AAF) and provide PKI and shibboleth capabilities.

The predominant cost in implementing the AAF involves harmonising policies and technical infrastructures across the sector, which will therefore require investment by participating institutions. Fortunately there are good reasons for institutions to re-work their authorisation systems for their own business reasons and so the AAF can be treated as an integration requirement on institutional investments rather than as an add-on.

The other major cost involves the adaptation of applications to interface to the AAF, which again fortunately is under way for many applications; and has already been achieved for the basic web server technology used world wide.

Like the AREN, the investment plan has adopted a preference for the ongoing operation costs to be shared amongst the institutions that gain value from them, as the services should deliver real cost savings into those institutions. The plan allows some support for the start-up phase to assist uptake and develop services in (2008-2011).

Data Management Infrastructure

The upper bound on spending is very large if consideration were given to providing a national infrastructure, as in the case of the AREN, that might hold all research data of national significance. The investment plan has been compelled to place this possibility into the 'landmark infrastructure' class and therefore out of scope.

Instead it has focussed on missing components within the infrastructure that if provided would support the effective re-use of research data on a national basis. Therefore 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 stewardship for some data collections will be supported, using capability around existing research data retention facilities, and aiming at a leadership example of research data management.

The investments will 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
- A merit scheme for prioritising the best use of the limited resources.

Unfortunately an agreement has not yet been achieved within the community for how and how various participants might deliver these services. In addition, a large number of investments are currently being considered in this area. Consequently the implementation decisions relating to this area are likely to be made in three or four months time.

National Computational Infrastructure

The need for High Performance Computing (HPC) is identified across a wide range of disciplines and research fields, and is required by research activities 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 national expertise in HPC.

The spread of e-Research is also expected to be accelerated by the propagation of tools and analysis systems (via portals and similar encapsulations).

Therefore the national computing infrastructure will provide those computational services which can best be funded and justified at the national level, including:

- Advisory and support expertise to assist computational infrastructure investments
- An Australian capability system commensurate with international practice, through the continuation of the APAC National Facility at similar levels to past investment
- Establishment of application specific regional shoulder systems
- A community based merit allocation scheme

The establishment of regional and specialist systems is aimed at achieving application benefits not possible by simply growing generic services and to involve state and local institutional investment. Cases could be made for geo-sciences, physics, energy, bio-informatics, climate, water, eco-systems, characterisation, and perhaps general engineering and general data mining. AeRIC will develop a process that selects disciplines and sites for the three systems proposed within the funds available from NCRIS. Overall the investment in these three systems is expected to be about 25% of the funding available and the national facility is expected to be about 75%.

Interoperation and Collaboration Infrastructure

For e-Research to become easier for new e-Researchers, the complexity of the underpinning IT systems needs to be hidden. This investment creates an Australia-wide inter-operation platform linking expertise with compute and data support systems across regional and national service providers into a day-to-day national operational activity.

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.

The functions provided 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

- The day-to-day operations management of the above services.

Given that the APAC National Grid is an already operating national inter-operation infrastructure and houses a large portion of our expertise in this activity, the operators of that infrastructure have been asked to generalise their current services with an enhanced focus on data and web services.

The new investment specifically re-purposes the expertise developed in the APAC National Grid to focus on integration of resources not within the current APAC members.

National E-research Architecture Taskforce (NEAT)

Early in the planning an investment initially known as “tools” and later recast as “Application and Discipline Services” was developed to address the problem of supporting new e-Researchers; a problem which was thought to be well connected to the problem of developing the next generation of services in the infrastructure.

The expertise to do that is of course not separable from the other investments and so this activity is really about bringing a focus across expertise and organisations to develop new services relevant to identified research collaboration needs.

The plan proposes the establishment of a National E-research Architecture Taskforce to involve Australia’s leading practitioners in collaboration technologies and services to identify and select suitable activities.

NEAT will have about \$8M of effort per annum available to focus on delivering e-Research benefits into priority research communities or new core services. NEAT will also provide the world class technical expertise and advice AeRIC will need to develop a sustainable evolving e-Research fabric that can support increased e-Research adoption.

Specifically NEAT is intended to identify, scope and propose activities to AeRIC that include:

- Projects that ‘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
- The hardening of new services in preparation for their on-going operation.

Summary and Proposed Funding Levels

The following table summarises the investment components and proposed NCRIS and contributed resource levels expressed as totals across the four years included within the plan (2007-2011).

Component	Arrangements and relationship to foundation principles	NCRIS	Others
AeRIC	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. The Executive Director will lead component cohesion, undertake component reviews, and develop an annual business plan.	2	0
NEAT	The National e-Research Architecture Taskforce (NEAT) will advise on the services and components that need to be made operational as part of the evolving national e-Research infrastructure. Projects sponsored by NEAT will build on more operational services to better meet needs from discipline and NCRIS capabilities. The expenditure is attributed to the ANDS and ICI entries below.	(11)	(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; provide custodial and analysis services for some data collections; and assist researchers and institutional and develop towards best practice.	21 ¹	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 operates with an internationally recognised quality of service. Other investments will support specialist computing where such additional resource can be identified as a critical need to leverage the impact of other NCRIS investments.	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.	20	12
AREN	AARNet provides Australia's national research backbone and the AREN service. 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 and operate PKI and shibboleth services required for collaboration. 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

¹ An implementation plan for ANDS has yet to emerge, these estimates are approximations only.