

A PILOT IN E-RESEARCH SUPPORT DISCUSSION PAPER

Executive Summary

Globally there are many activities in e-Research fueled by large investments from various governments, engendering much enthusiasm within segments of the research community. There are big science questions that can be posed and solved because the underlying existing compute/data/network infrastructure can be better utilized. Australia has most of the components required for engendering e-Research installed, thanks to recent government initiatives. However, it appears that a vital component is missing: incentives for researchers to engage in e-Research, to grid enable their existing software, to develop further software to exploit the installed compute/network/data infrastructure, and incentives to collaborate in multi-disciplinary projects. This discussion paper outlines a number of initiatives, including a series of capital city consultative workshops to take place in June 2004, and a tentative pilot e-Research support program, intending to engender discussions in the higher education sector on how to best provide such incentives.

Background

The term e-Research (e-Science in the UK context) is used to encapsulate those activities based on the invention and exploitation of advanced information and communications technologies. These activities include:

- Generating and analysing research data:
 - from experiments, observations and simulations,
 - providing quality management, preservation and reliable evidence;
- Developing and exploring models and simulations:
 - enabling computation and data at extreme scales,
 - providing trustworthy, economic, timely and relevant results; and
- Enabling dynamically distributed virtual organizations:
 - facilitating collaboration within the research sector and with industry through information and resource sharing;
 - providing security, reliability, accountability, manageability and flexibility.

Overseas, there are many activities in e-Research programs. Table 1 gives an indication of the investment by respective countries or regions in e-Research.

Table 1: The investment in grid technology overseas. The figures shown are in Australian dollars.

Country / Region	Name of program(s)	Expenditure
USA	Advanced Cyberinfrastructure NSF Middleware Initiative	\$1.3 Billion over 3 years
United Kingdom	UK e-Science Programme	\$450M over 5 years
European Union	EGEE, GÉANT (FP6), etc	\$486M over 4 years
Japan	Information Technology Based Laboratory (ITBL) National Research Grid (NAREGI) Grid Technology Research Centre (GTRC)	\$1billion over 5 years

These investments engender many activities in e-Research overseas. There are a number of components in an e-Research program. They include:

- Better utilisation of existing infrastructure, without incurring a higher cost in installation of capacity;
- Better utilisation of collected databases;
- Collaboration between researchers to solve scientific problems, in particular domains, which could not have been solved previously due to lack of available resources; and
- Collaboration between researchers and computer scientists to solve fundamental problems in their respective domains.

Globally, due to these investments worldwide, we have witnessed the following:

- Better utilisation of existing infrastructure – For example, in the UK, there are activities in using existing computer servers and network resources, connected together to solve some fundamental issues in particle physics, which could not have been solved previously. The data processing of the vast amount of data obtained from the Large Hadron Collider (LHC) experiments at *Conseil Européen pour la Recherche Nucléaire* (CERN) led to the possibilities of processing much larger volumes of data in a shorter time, without significant increase in the investment of basic infrastructure.
- Nascent activities in the UK e-Science program on data curation – This is an activity designed to select and archive collected databases, and then make them available to other researchers.
- A much closer working relationship between computer scientists and researchers in a particular domain to solve fundamental problems, which could not have been tackled previously – For example, eDiaMonD is a research program for processing digital mammography. Because of the amount of data obtained from digital mammography, bio-medical and medical researchers are collaborating with computer scientists to solve the information processing problems. As another example, the myGrid project is designed to provide a workbench for a bio-informatician, so that the types of information which a bio-informatician needs to enter into his/her log book using various bio-informatics databases or data services in the search of matches for DNA sequences, can be captured automatically and archived for later retrieval. This project engenders a close collaboration between a group of computer scientists and a group of bio-informaticians. The result is that bio-informaticians are able to trace what they have done, and have the possibility of going back and re-running any segment of procedures, which they had performed previously. From the computer scientists' point of view, they perform some fundamental research in workflow management and workflow optimisation, using the semantic grid technique.
- Collaboration of researchers and industry partners to solve problems – This could not be tackled previously. For example, the Distributed Aircraft Maintenance Environment (DAME) project is a collaboration between a number of partners, including the University of Oxford, University of York, University of Sheffield, and Rolls-Royce. The aim of the project is to monitor the performance of aircraft engines while they are in flight in real time so that when the aircraft arrives at a destination, if necessary, the component parts can be available to replace deficient or faulty engine parts. This can potentially transform Rolls-Royce from an aircraft engine manufacturing company to an information service company, providing a whole-of-life service to the aviation industry. Such collaboration is made possible through the availability of the network and computer infrastructure, and the incentives provided from the UK e-Science initiative.

In Australia we have most of the components in place for e-Research. These include:

- **Computer infrastructure** – There are a number of high performance computers which are connected to the network and which have enough capacity for allocation to users. These facilities include: large computer clusters in Canberra, managed by Australian Partnership for Advanced Computing (APAC); in Melbourne, managed by Victorian Partnership for Advanced Computing (VPAC); in Sydney, managed by Australian Centre for Advanced Computing and Communications (ac3); in Brisbane, managed by Queensland Parallel and Supercomputer Facilities (QPSF); and a number of smaller computer server installations elsewhere.
- **Network infrastructure** – Because of the investment by the Australian Government through Australian Research and Education Network (AREN), there will soon be: 10 gigabits per second (Gbps) bandwidth among the metropolitan cities, 1 Gbps to most universities' main campuses, and a 10 Gbps link across the Pacific interconnecting with other high-speed academic and research networks in the US, and through them with Europe and Japan.
- **Data repository** – There are large data storage facilities available such as APAC/ANU in Canberra, ac3 in Sydney, and QPSF in Brisbane.

As yet, there is no installation of software which enables users to utilise these existing computer/data resources in a seamless and transparent fashion.

ARC Activities in e-Research

- A high-level delegation recently participated in an e-Science mission to the UK, led by Dr Mike Sargent and organised by the Australian Academy of Technological Sciences and Engineering (ATSE). Professor Tsoi, ARC Executive Director, Mathematics, Information and Communication Sciences, participated in the mission.
- The ARC has undertaken a number of initiatives in the past two years to promote e-Research in Australia. The ARC's Strategic Plan aims to ensure that Australia is a participant in the next generation of international e-Research developments, such as grid computing.
- The ARC, through its Special Research Initiatives program, has already supported two e-Research initiatives: an 'N to N' meeting with key players of the UK e-Science program; and support for Dr Anne Trefethen, Deputy Director of the UK e-Science program, to visit Australia. The N to N meeting created considerable interest in e-Research for those who participated in the meeting. Dr Trefethen discussed e-Science programs in the UK with a number of groups, including the High Energy Physics group at Melbourne University, the Australian National University, the Australian Research Information Infrastructure Committee (set up by DEST under the Systemic Infrastructure Initiative), the University of Queensland, and Queensland University of Technology. The ARC is also jointly funding a UK mission, with Australian Partnership for Advanced Computing (APAC), for younger researchers to take place in late May 2004.
- Professor Tsoi addressed a recent annual general meeting of the Council of Australian University Directors of Information Technology, and described the role which University Information Technology Services can play in facilitating the development of e-Research.

Support for e-Research in Australia

In general, there is support for an e-Science program in Australia.

There is, however, a general accepted perception that to establish an e-Research program in Australia, we will need to address:

- Installation of a common set of core middleware so that existing computer/network facilities can be used by researchers;
- User and system support for researchers; and
- Incentives for researchers to grid enable their existing programs, to develop further middleware components based on the core middleware, and to initiate research collaboration with computer scientists to further enhance the core middleware.

APAC has said that it plans to install the core middleware in the very near future, which will enable a remote user to make use of the network capacity and the computer capacity to perform computations, which could utilise these resources in a seamless and transparent fashion. In addition, APAC may provide limited user and system support for researchers. In any case, the incentives for researchers to grid enable their existing software is not necessarily predicated on the use of APAC facilities, as there are other infrastructure providers who could provide the needed computer/data infrastructure. The APAC facility is a national facility, open to all users in the higher educational sector, and in normal circumstances, these would be the facilities that researchers are expected to use. However, researchers may, if they wish, use other facilities such as the ones existing in international collaborations.

The Australian Academic and Research Network (AARNet) is exploring ways in which the international bandwidth to the US can be better utilised to facilitate e-Research activities and collaboration among Australian and US research groups.

Macquarie University through an ARC Special Research Initiatives program is organising a workshop in Sydney on 10 June 2004, to bring a number of key players in e-Research together to consider ways in which the above issues on infrastructure can be overcome. The participants in the workshop will include representatives from: APAC, AARNet, GrangeNet, DCITA, DEST, and Dr Sargent (ATSE). There will also be infrastructure service provider representatives, and user representatives from various universities and potential application providers such as the Distributed Systems Technology Co-operative Centre (DSTC CRC).

The Proposed Pilot

It is proposed that the ARC, through its Special Research Initiatives program, would provide funds for a pilot in e-Research support to facilitate the development in providing the needed incentives for researchers to make use of these resources. The aims of this program are to provide incentives for researchers to utilise the installed infrastructure, and to collaborate in e-Research projects. The objectives would be:

- To encourage greater collaboration among Australian researchers to make use of the existing computer/network/data infrastructure;
- To encourage the development of international research linkages through the use of common computer/network/data infrastructure;
- To encourage the co-operative development of high-quality research capacity in emerging areas; and
- To enhance the scale and focus of research in Priority Areas.

The ARC plans to run a series of capital city workshops in June 2004 to consult with the higher education sector on the best ways forward in providing incentives for researchers to engage in e-Research. It is expected that the consultation with the sector, together with the outcome of the Macquarie organised e-Research workshop in June, will help shape the requirements of the proposed Special Research Initiatives program of e-Research support.

The ARC is aware of major national initiatives in enhancing existing research infrastructure in Australia, for example, the recently announced National Collaborative Research Infrastructure Initiative indicated in the “Final Report of the National Research Infrastructure Taskforce” released in March 2004. The ARC will participate fully in these initiatives and their coordination. The e-Research support initiative indicated in this discussion paper is a pilot activity intended to engage the research community in e-Research.