

# eResearch Australasian Conference Report

**Brisbane June 26-29 2007**

**David Thorns**

The conference was the first one to bring together the emerging eResearch community across Australia and New Zealand. It grew out of an appreciation of the need to link the various groups that contribute to the development of eResearch and was attended by 315 people drawn from 80 different organisations.

A large range of issues were addressed in the 48 papers presented at the conference covering such issues as the development of eSecurity and the formation of the Australian Access Federation, Middleware development, sustainable repositories, and new collaboration tools and strategies and the development of new ways of doing science in the 21<sup>st</sup> century.

The full range of presentations can be accessed at <http://www.eresearch.edu.au/program>

## Issues that emerged

### **New paradigm for science**

A consistent argument through many of the papers was that a new paradigm in science was emerging, one where data was the key driver (de Roure). This arises from the greater amount that is now available and the increasing processing capacity and connectivity that is provided by HPCs and Advanced Network and new computational and collaborative tools.

One question this raises is how this new paradigm enables the exploration of new research agendas across disciplines that have previously not been engaged in debate. Is there also a need for new languages of communication? This is especially important if we are to bridge the natural sciences, social science and humanities to address the significant global research agendas that include climate change, sustainability, biomedical and genetics, environmental concerns and energy.

The characteristics of the emerging science paradigm identified are that it is data driven, global in scale, collaborative in practice, underpinned by High Speed computers and hi speed fibre networks to allow real time processing and collaboration.

Two further changes flow from this new landscape of research activity. One is the need for robust and accessible data archiving. This in turn creates the need for storage which is now in terabytes, which raises questions about how and who will take responsibility for the support, cost and management of such repositories. Further as the science is global it needs to be part of network of repositories which are accessible from both local and remote locations. Coordination and collaboration across and within disciplines will be required to ensure the standardisation of such data systems to ensure that open access is achieved.

It was noted that "open access" did not necessarily imply "free" access but the access needed to be costed in such a way it is not a deterrent to scientific activity. Here there were strong argument put resisting volume charges on the basis that

these were administratively costly and deter use and thus could limit scientific use and development. The new model of activity also has implications for publishing with a shift to electronic publishing and a greater integration between the journal text and the data that has been responsible for the work which is archived (Bourne).

### **Collaboration**

The central importance of collaborative research with the new paradigm was stressed in many of the papers. The nature of the collaboration was also explored in a number of the sessions and this showed there are a number of aspects to this development the most crucial one is the relationship between people and technologies.

Networks are about connecting HPC clusters to enable the combined power to be used for high end data analysis – life science and for visualisation applications – global networks now required to enable the data to be exchanged and to provide the necessary computational power.

In this new matrix of resources and network connectivity KAREN is a very significant development to enable New Zealand researchers to both contribute and benefit from the growth in the new areas of research. A small country, well connected can be a global player and can be a collaborator in projects and programmes that without the Hi Speed network we would be excluded. This is increasingly the case across all areas of science and knowledge creation.

The three key users of advanced networks were seen as the entertainment and media industries, the military and security and enforcement (police and border security) and the research and educational communities. Collaboration is being driven by technology needs as the solution to researchers problems now requires increase computing power computing so collaboration between machines and networks is essential. Within this though there is still the important place of the shaping of the research questions and this is where the creation of Virtual Research Environments and people to people communications system are vital (Searle and Jones).

Collaboration needs to be both horizontal and vertical and the silos that have characterised some areas of knowledge creation need to be broken down. This is where the push for greater interdisciplinary is generated. Some of the papers saw this as creating more dialogue between sub sets of disciplines, for others it was the linking of natural sciences and computer science and IT technologies. A further push is to the linkages across the natural and social science and humanities as the global research agenda is explored within a more connected environment.

What is the nature of the interdisciplinary being promoted/encourage/engaged?  
The experience within the Building Research Capability in the Social Sciences programme (BRCSS) is that the “low hanging fruit” of the new collaborative system of virtual laboratories and Advanced Video Conferences (via for example Access Grids) are an important step towards building virtual research environments (Pailthorpe). These arise when there is trust built between researchers and recognition of the need to collaborate to address increasing transdisciplinary research agendas. Advanced Video Conferencing systems enabled by advanced networks, also have storage requirement for archiving and analysis when used as research development and analysis tools.

Thus there is a need to reliable, easy to use, well maintained and managed systems that are as close as possible to f2f environments. Collaboration needs to be understood as a complex relationship between people/agents and is a new way of both working, sharing “space” and combining elements of the “real” and the virtual” (Thorns and Allan)

## **Data**

The new model of science is data driven as an exponential growth in data was acknowledged. This is particularly the case in such areas as astronomy and life sciences. In the former this doubling every 6 months creating significant storage and computation issues (Szalay). In life science the use of high intensive Xrays produced by the third generation light source (synchrotron) providing new visualisation of materials and coupling science and computing in new ways (Dimper).

These create significant data volumes that require storage and curation and raise the question of who should store and for how long and whether the new data replaces the old or adds to the stock. A number of times the intriguing question was raised as to how long data needs to be stored. Researchers generally, regardless of discipline, are reluctant to dispose of data.

Storage and user interfaces and their maintenance with distributed systems and seamless connections including identity management and access protocols are significant items of the global research agenda. The division here between roles of funders, providers and users are still being debated.

In Australia as part of its 5 year Road Map for eResearch infrastructure a there is planned the formation of the Australian National Data Service (ANDS). The creation and establishment of this was acknowledged as one of the most difficult part of the infrastructure to establish (Francis and Lansdown).

For the Social Sciences the new paradigm of data driven science is more problematic as the data that we use is more variable, not always digitised and more difficult to standardise. Collecting consistent data across different societies and cultures poses problems of interpretation and draws attention to the importance in social science of context and time, data is collected from different communities who may have varied understandings and degrees of literacy. Data in the social sciences consists of:

1. Regular national and international Surveys, these are the types of data that have been mainly stored in the major social science archives such as the ESRC data archive in the UK and the International Political and Social science archive at the University of Michigan. The data held in these is predominantly quantitative.
2. National Statistics agencies that collect regular statistics and carry out regular surveys on behalf of governments. This data is the source of many international comparative analyses and the creation of league tables of national and international activity by such organisation as the OCED, UN, and World Bank. This data is widely used and the use of advanced network and computational resources aids in the national and international exchange and use of this data. Some of the critical issue here though are again about the nature of the data and the measures used and definitions employed by the various national agencies.
3. Administrative Data which is held by Government Ministries and agencies. This is seen as a rich potential source for research use and analysis as it is

4. Ad Hoc Surveys these are usually unique surveys which are hard to track and collect without a requirement for their storage by the major funding organisations of research.
5. Qualitative data the storage and retrieval and reuse of this type of data is complex as it requires much greater appreciation of the context in which it was collected and the ethical agreements under which it was obtained from the research participants.

New Zealand Social Sciences are not well placed with respect to data bases as there is not a central repository for social science data comparable with that available to researchers within other countries. The funding has not been available in the past nor has there been a commitment from funding agencies to support the storage and curation of social science data. A new attempt to create such an archive has been supported by the recent TEC grant. However this is only for one year and still does not represent a long term commitment to creating this vital resource if social science is to make full use of the tools that eResearch is now providing

Social Science is a broad category that includes many disciplines some of which are likely to embrace eResearch before others. Illustration of social science at the conference were drawn from Archaeology and GIS (Geography). Use by Humanities also points to how the new tools can be used across disciplines not traditionally thought to be computer based. Here the uses were related to text mining programmes linking with bibliographic collections, the creation of digital text and access (Kilner), for example to the Australian Biography (Arthur)– where digital and on-line access was seen as replacing print as the source of authorial text.

A further use was the addition of wiki based facilities to allow annotation on collections and resources. Here an interesting example was of the Australian Picture archive (Hunter) which allowed multiple comments and annotations by different communities and individuals to be added from those who knew the person through to members of the research community. These facilities allowed a broader and more extensive discussion about the picture/object/text and the collaboration around the interpretations generated.

### **Capability and Training**

Considerable emphasis was placed upon identifying early adopters and exemplar projects to encourage the take up and spread of eResearch (Borda). The speed of change and the challenges to way we work were seen as requiring the development of new skills without which the growth of the use of the new technologies would be slowed. There was a need therefore for programmes of training, incentives to encourage people to invest in time and energy to develop the skills and middleware needed and train people in its use.

A strategic plan or road map was an important feature in this development and a number of significant ones have been recently produced – Australian was presented at the conference (Francis) and the European equivalent was released in April by ESFRI. The New Zealand KAREN roadmap has also been developed and due for

release in late July. All of this documentation points to the need to extend global collaboration as “no one can afford to go it alone” either in terms of cost or in terms of having the necessary capability and capacity.

Further essential is a co-ordinated approach that is not either “over managed” or over engineered but is create to facilitate and enable the development of research within the 21<sup>st</sup> century. This will include ways to transfer to whole research community and ultimately mainstreaming eResearch. Moreover strategic plans for development need to be inclusive across disciplines encouraging the growth of both trans and inter disciplinary activity to address emerging research questions and policy issues.

Papers referenced all available at; <http://www.eresearch.edu.au/program>

- David de Roure: eScience is about Scientists too
- Phil Bourne: Thoughts on the future of scientific dissemination  
Nick Jones and Sam Searle: BeSTGRID and KAREN: Collaboration capability for eResearch in New Zealand
- Bernard Pailthorpe: Collaborative working - from HPC, Vis to AG developments in Australia
- David Thorns and Mary Allan: Access Grid, video conferencing, and real life simulation
- Alex Szalay: Science in an Exponential World
- Rudolf Dimper: High performance computing for synchrotron radiation research
- Rhys Francis: Platforms for collaboration
- Anne-Marie Lansdown: All aboard, destination: Seamless
- Kerry Kilner: The Resource for Australian Literature
- Paul Arthur: Going digital: Humanities and the eResearch revolution
- Ann Borda: Expanded uptake and sustainable communities of use: JISC's role in shaping the UK e-Infrastructure for Research
- Jane Hunter: Harvesting community tags and annotations to augment institutional repository metadata
- European Forum for Research Infrastructures(ESFRI) European Roadmap 2006 (available at <ftp://ftp.cordis.europa.eu>)