Supporting E-Research at the University of Melbourne

Eve Young  
Project Leader  
Digital Asset Management Systems Implementation  
University of Melbourne  
e.young@unimelb.edu.au

Lynne Horwood  
Research Consultant, School of Graduate Studies  
University of Melbourne  
lynnemh@unimelb.edu.au

Shirley Sullivan  
Electronic Information Coordinator  
University of Melbourne  
s.sullivan@unimelb.edu.au

Abstract:  
E-Research refers to large-scale, distributed, information-intensive forms of inquiry conducted collaboratively between institutions, and intra- and inter-nationally. The paper details the strategies that the Information Division (ID) is implementing to support e-research and information management within the University. The culture at the University of Melbourne emphasises a client-focused approach where the ID staff partner with academic departments and business units to assemble suitable project teams to translate the vision into technical solutions. The paper will explore ID initiatives which support e-research and information management using new technologies which emphasise interoperability and flexibility.
Introduction

According to the E-Research Co-ordinating Committee established by the Australian Government in May 2005, "'e-Research' embraces research methodologies emerging from the increasing access to distributed high-performance computing resources, data resources, research instruments and facilities utilising the grid, Internet and local area networks and advanced communications technologies.” (Australia E-Research Coordinating Committee 2005)

E-Research is a collective term for the various initiatives on eScience, e-Social Science, e-Humanities, and cyberinfrastructure. E-research refers to large-scale, distributed, information-intensive forms of inquiry conducted collaboratively between institutions, and intra- and inter-nationally. The overall aim is “to do faster, better, and different interdisciplinary research” as summed up by Tony Hey, head of the U.K. e-Science programs. (Borgman 2005)

Some examples of e-research activities include generating and analysing research data from experiments and observations, and developing and exploring models and simulations. A recent example is a University of Melbourne-led project which will enable researchers from a number of Australian institutions to access distributed datasets that can be integrated, searched and queried by researchers in each institution. The project follows a successful pilot which managed the security of clinical data, research records and genetic data as well as addressing privacy, intellectual property and ethical issues. This project highlights many of the issues to consider when building collaborative e-research environments. These include technical infrastructure, standards, interoperability, security, middleware, sharing of and access to resources between institutions and subject disciplines, licensing, user and community needs, curation, preservation, and digital rights management. (University of Melbourne Media and Publication Services Office 2005)

Grid

The Australian Government established its e-Research Co-ordinating Committee to coordinate e-research nationwide, in recognition that, “exploitation of ICT now underpins all research disciplines, including the social sciences and humanities.” (Australia. Minister for Communications, Information Technology and the Arts 2005) The Grid is a collection of distributed computing resources (data repositories, specialised scientific equipment, computing power, knowledge services) that enables researchers at spatially remote locations to work together interactively. (O’Brien and Young 2005) Generic e-research comprises three elements: Computational Grid (primarily designed to support high performance computation), Access Grid (designed to support collaborative visualisation involving distant researchers) and Data Grid (designed for transporting large volumes of data). (Harris 2005)

These resources will be individually owned and managed by various organisations; however, the Grid infrastructure will enable collaborative use of these resources globally. The major problems being addressed by Grid developments are the social problems involved in collaborative research.

The Grid community has developed the Access Grid, which enables flexible, secure, scaleable, coordinated, interactive resource sharing among individuals, institutions and
resources. It can be used for remote control of instruments, group-based video conferencing, visualisation, seminars, lectures, training and other collaborative work, whether from dedicated rooms, lecture theatres, portable nodes, or the desktop. (Fraser 2005) Typical features of an Access Grid node are a very large display to show many participants simultaneously, life-size images and shared applications; natural, hands-free, full-duplex audio; and efficient use of bandwidth employing multicast networking. (Daw 2005)

In multi-disciplinary collaborations, teams need to invest time in learning to understand each other and develop a shared understanding of scientific paradigms, modes of researching, and theoretical perspectives, in addition to understanding the meaning of the data available, and of the research questions that could be addressed. Semantic interoperability is not just about data, but about a much broader social context that provides meaning to data, its analysis, and its interpretation. (Craglia 2005)

**Legal and social dimensions**

The Internet enables data movement across administrative, legal and national jurisdictions as easily as it can be moved to the next desktop. Grid technology includes the possibility of sharing resources and data at an international level, and different ownership regimes between countries have to be accounted for. (Hinds 2005) Data sharing and transmission across disciplinary and national boundaries raise issues of privacy, security, access rights and liability, as well as commercial or complex intellectual property rights considerations. Successful distribution of data on a national or international level requires clarity about the rights and responsibilities of those who provide and use these. Middleware enables researchers to share data while maintaining security and confidentiality. (Guy 2005)

Grid technology, with its seamless access to multiple datasets, multiple users, and increased availability of computational power, carries an increased risk of both accidental disclosure and disclosure as a result of deliberate attack. (Elliot 2005) The challenge is to use information derived from data without breaching confidentiality. Breaches of confidentiality that result from the actions of one researcher can threaten the ability of scientists everywhere to collect and use data. Preserving access to high quality scientific data is essential to the empirical replication that is at the core of good science. (Burton 2005)

Establishing a clear and definitive legal position in relation to the ownership of data will greatly enhance the opportunity for academic research and industrial innovations. (Hinds 2005) It may be critical in preventing expensive litigation. Researchers must be permitted to protect their work, and industrial partners allowed to establish fair licensing arrangements. The definition of legal structures for the ownership of intellectual property rights will play a vital role both in clarifying concerns and enforcing compliance. (Hinds 2005) Researchers may not wish to contribute their hard won data to the common pool until they have published their research results. (Hartswood 2005) Tensions which may arise over sharing of data may be resolved at a contractual level. For example, a clinician's research data may be licensed for use by another research group, but only on terms that ensure the resulting work will not compete with the clinician's own research. (Hind 2005)

With its establishment of the e-Research Coordinating Committee the Australian Government acknowledged that issues of common interest should be addressed in a more cost effective way by coordination and collaboration, nationally and internationally. The Committee will establish strategy and policy to co-ordinate support for research, utilising advanced research
infrastructure, including distributed high performance computing and research data resources and high speed networks. (Australia. Minister for Communications, Information Technology and the Arts 2005)

Recognising the importance of successful sharing of research, the Australian Government funded MAMS (Meta Access Management System) in 2003. MAMS supports authentication, authorisation of identities, digital rights, search services and metadata management. In a similar initiative in 2005, the Government funded the Middleware Action Plan and Strategy (MAPS). The MAPS project will build a strategic plan of activities and projects for an Australian collaborative middleware strategy. Partners include Council of Australian Directors of Information Technology (CAUDIT), Council of Australian University Librarians (CAUL), and AARNet. (Australia. Department of Education, Science and Training 2005)

In August 2005, the Minister for Education, Science and Training announced funding for nine projects known collectively as the Managed Environments for Research Repository Infrastructure (MERRI) Projects. The MERRI Projects will help Australian researchers to better manage outputs of e-research, such as medical and financial data, and store them in reliable repositories and assist them to access these rich data sources across institutions, building greater opportunities for collaboration and maximising Australia’s research output. (Australia. Department of Education, Science and Training 2005)

The E–Security Framework for Research seeks to establish a framework which will integrate two different types of security systems, PKI and Shibboleth, to foster collaboration and enable the secure sharing of resources and research infrastructure within Australia and with international partners. Project partners include Macquarie University, Council of Australian Directors of Information Technology (CAUDIT), the Australian Partnership for Advanced Computing (APAC) and AARNet. (Australia. Department of Education, Science and Training 2005)

Curation

E-science is a new paradigm of research, often characterised by a deluge of data analysed by massive distributed computing power. Researchers generate vast amounts of raw digital research data, but current digital technology has a shelf life of 5-10 years. The scientific record and the documentary heritage created in digital form are at risk from technology obsolescence and by the fragility of digital media. Research funding bodies such as the US National Institutes of Health (NIH) are requiring that applications for research projects include plans for management of data resulting from collection, creation and analysis. Preserving the 'project', comprising data, publications, workflows and the 'grey' material of reports, notebooks and other forms of more nebulous communications, is important in a research environment. (Fraser 2005) Other forms of research output, such as X-rays, laboratory results, pathology slides, reports of experiments, research datasets, and software also require data curation. (Weaver 2005). There is increasing integration of text-based information with its underlying data. Digital libraries can form part of the infrastructure to support research in scholarly disciplines. Library staff have expertise in managing, making accessible and preserving scholarly data and they collaborate with information technology professionals and researchers to make scholarly works accessible online to the broader research community. (O’Brien and Young 2005)
Data curation is not only about long-term care of artefacts. It is also about maintaining their integrity and enabling and promoting their availability to appropriate audiences. Effective curation means that data need only be collected once and can be re-used or enhanced in the generation of new research. Curation may also be a legal requirement, or essential to validate published research results. (Van der Knijff 2005)

The Australian Government financially supported data curation in its August 2005 round of funding. The University of Melbourne is a partner in the BlueNet project which will provide a highly distributed archiving facility to support the long term data curation requirements of Australia’s marine science researchers. The project will link vast data repositories and marine resources that currently reside in individual academic institutions with government institutions in Australia and overseas. Three quarters of Australia’s marine science data is contained in individual universities and not accessible to the wider researcher community. BlueNet will build infrastructure to enable the discovery, access and online integration of multi-disciplinary marine science data on a very large scale to support current and future marine science and climate change research, ecosystem management and government decision making. (Australia. Department of Education, Science and Training 2005)

The Dataset Acquisition, Accessibility and Annotation e-Research Technology Project (DART) will provide solutions to deal with the life cycle of research from lab book to formal outputs to teaching. It will draw on differing needs and discipline perspectives of a range of research communities by pulling together expertise in marine science and sensor Grids, expertise in distributed systems, database research and large data set management, and Grid computing research. (Australia. Department of Education, Science and Training 2005)

Data repositories

Data that used to be collected and made available through government agencies has become privatised over the past two decades. In the past few years, however, the pendulum has swung back to the belief that tax-funded research results should be made freely accessible to taxpayers. Research funders, such as the Wellcome Trust, are requiring open publication of research results and their underlying data as a condition of funding. The Research Councils United Kingdom (RCUK) will make it a condition of grant funding that research output be freely accessible online to give British research greater worldwide impact. Governments are also recognising the value of free access to large data sets in promoting further research. In the US, the NIH supports several freely accessible databases, such as Genbank and PubChem. Open access journal platforms are linking to such data repositories. Publications produced by the Australian Bureau of Statistics are now freely available from its website, after several years of subscription-only access, and the Australian Government has also made the Cochrane Library available to all citizens. The University of Melbourne is a partner in a Government funded project, Legal Protocols for Copyright Management: Facilitating Open Access to Research at the National and International Levels. This project will develop a set of legal protocols and generic licences that can be used across universities to break down barriers to open access to copyright material. The Legal Protocols project will also develop best practice guides for managing copyright issues in open access environments and a rights expression language that can be used to technologically enhance open access to existing and proposed electronic stores of research and other data. (Australia. Department of Education, Science and Training 2005)
University of Melbourne and e-research

Research repositories providing deposit and access services for research outputs are a key component of an institution's e-infrastructure. (Fraser 2005) Stored data is worthless if it cannot be found, and navigating data is difficult if the data is unstructured, inconsistent and plentiful. Metadata must be associated with stored data to ensure it can be accessed and understood. (Fraser 2005) APSR (Australian Partnerships for Sustainable Repositories) aims to establish a centre of excellence for the management of scholarly assets in digital format. Its overall focus is on the critical issues of access continuity and sustainability of digital collections within partner institutions. The work of ASPR has international as well as national and institutional dimensions. (Coleman 2005)

University of Melbourne has become an APSR test bed site, focusing on sustainable standards-based management and storage of research data and raw research output. Melbourne’s role includes establishment of an e-Research Expertise Centre. This will build on existing local, national and international expertise, tools, software, and information infrastructure to develop an integrated and cross-disciplinary model for exchanging information between research groups. The Centre will also support the national e-research community in managing and sharing research output, specifically large sets of data and information generated in the course of research.

The relationship between research and information technology needs to be reflected in organisational infrastructure and services. The Information Division’s Information Management Plan stresses the need to “establish local, national and international partnerships and develop a suite of services and infrastructure to position the University as an international leader in e-research.” The Information Division is developing a support service for University of Melbourne researchers and e-research initiatives, to connect them with new and emerging software tools and expertise. Continual improvement of the information infrastructure is critical to ensure projects are developed with new technologies and devising innovative ways of using existing technologies. (University of Melbourne. Information Strategy Committee 2005)

The Information Division can offer leadership and coordination to academic researchers by brokering their needs. University of Melbourne researchers are collaborating with colleagues and industry experts to develop further e-research projects. In response to an Australian Research Council discussion paper (Australian Research Council 2004), an e-Research Working Group was formed in 2004. The Working Group sought proposals from a broad range of research areas across the University. Information Division staff are partnering with research groups preparing grant applications. This collaboration can produce excellent results; for example, the University received a hundred percent hit rate in the Australian Research Council Special Research Initiatives funding round announced in August 2005. Current initiatives include developing a network of experts from the faculties and forming a steering committee to support e-research. Other examples are providing connections to support structures or expertise on campus or to external collaborators. The Information Division brings to the University research community lessons learned from its participation as a lead partner in the ARC funded Australian Digital Theses Expansion and Redevelopment.

The University is developing a comprehensive cross-disciplinary approach to Grid computing in order to advance e-research. Initial work will include a University wide assessment of
potential directions for engagement and further discussions with other universities about developing common programs and strategies. The Access Grid makes use of high bandwidth internet connections, data compression algorithms and fast data processing to provide a rich set of tools using video conferencing and data sharing. This enhances the ability of our researchers and students to better access data resources of all types, whatever the provenance of the data. (University of Melbourne. e-Research Working Group 2005)

Many research groups have traditionally depended on local compute services. While this model is common and affordable because of decreased hardware costs and clustering capability, it can lead to duplication of effort, risks to systems management and security issues. The University of Melbourne Information Management Plan 2006-2008 proposes to establish a sustainable university-wide ICT shared service model with systems, policies and processes that enable collaborative work, best practice management and cost effectiveness. ICT infrastructure includes high performance computing, networks, data mining, compute clusters, render-farms, specialised test-machines, storage elements, and the computer visualisation facility.

**Conclusion**

The challenge is to create an infrastructure for researchers which will enhance their productivity and effectiveness. Researchers may be at very different stages in their awareness and use of new technologies. Legal and cultural differences may pose obstacles to collaboration. The University’s participation in national and international e-research projects requires a university wide strategy for e-research to ensure that organisational and technical factors are carefully considered and addressed. The Information Division can fulfil a lead role in expanding and enhancing research across the institution.
Bibliography


Craglia, M et al 2005, ‘Building bridges between social science, Grid, and geospatial communities: a reflection on practice’ in First International Conference on e-Social Science,
Manchester, UK 22-24 June 2005, National Centre for e-Social Science, Manchester, viewed 1 September 2005,
David, P 2005, Towards a cyberinfrastructure for enhanced scientific collaboration: providing its ‘soft’ foundations may be the hardest part, viewed 1 September 2005,

<http://www.ncess.ac.uk/events/conference/workshops/ncess2005_daw_paper.pdf>

Elliot, M, Purdham, K & Smith, D 2005, ‘Confidential data access via the Grid: an outline of the issues and possible solutions’, in First International Conference on e-Social Science, Manchester, UK 22-24 June 2005, National Centre for e-Social Science, Manchester, viewed 1 September 2005,


Hartswood, M et al 2005, ‘Etiquettes of data sharing in healthcare and healthcare research’, in First International Conference on e-Social Science, Manchester, UK 22-24 June 2005, National Centre for e-Social Science, Manchester, viewed 1 September 2005,

Harris, R 2005, ‘Towards an understanding of research ethics for computational, e-social science’ in First International Conference on e-Social Science, Manchester, UK 22-24 June 2005, National Centre for e-Social Science, Manchester, viewed 1 September 2005,

Hine, C 2005, ‘Material culture and the shaping of e-science’, in First International Conference on e-Social Science, Manchester, UK 22-24 June 2005, National Centre for e-Social Science, Manchester, viewed 1 September 2005,


Palmer, C 2005, ‘Scholarly work and the shaping of digital access’, *Journal of the American Society for Information Science and Technology*, vol.56, issue 11, pp. 1140-1153


---

**Endnotes**

i The Committee draws its membership from CAUL, ARC, AVCC, NHMRC and others and will be supported by a reference group, drawn from research institutions, research network providers, industry and funding agencies.

ii The e in escience is being used as an abbreviation of enhanced rather than electronic. (Atkins 2005)

iii [http://www.apsr.edu.au](http://www.apsr.edu.au)