

Mind The Gap: eResearch in Australia

Paul Davis, **Gabrielle Bright**

Victorian eResearch Strategic Initiative, Australia

Abstract

VeRSI - the Victorian eResearch Strategic Initiative - was funded in 2006 by the Victorian State Government. This paper will describe how the VeRSI program is providing a cohesive and coordinated approach to accelerating the uptake of eResearch by Victorian researchers. This is the first initiative of its kind in Australia.

VeRSI will deliver research leadership by harnessing enabling technology in a way that fosters a productive, collaborative research environment, and by developing key examples of how eResearch can enhance research outcomes. This paper will describe the five complementary projects that will establish essential infrastructure, develop resources for collaborative working and support exemplar use-cases and the development of applications in the Life Sciences.

VeRSI, like similar eResearch activities elsewhere, is acutely aware of the IT skills shortage developing in Australia. This exacerbates the need to develop skills and expertise in eResearch. This paper will highlight the need to bridge the skills and expertise 'gap' between research disciplines and technology implementations.

1. Challenges to research in the past

In the nineties, research communities in science and engineering began to tackle the so-called "Grand Challenges"; fundamental problems with broad application. The Grand Challenges were so large and complex that they could only be addressed by large teams using super-computers and would take decades to solve. Studies in computational fluid dynamics would enable the optimal design of vehicles, provide more accurate weather forecasts and improve efficiency in the oil industry. A computational approach to modelling the electronic structure of matter would deliver new chemical catalysts, drugs to boost the immune systems and novel superconductors. The physics underpinning fusion technology would lead to the energy sources of the future and bioinformatics and computational chemistry would provide the fundamental understanding of molecular systems necessary to cure cancer, diabetes, cardiovascular ailments and more.

A watershed in computer science provided the algorithmic tools needed for such endeavour and technology delivered phenomenal improvements in the power and speed of high performance computer systems. One by one the Grand Challenges were reduced to large but routine computations and even more exigent challenges filled their place; but the nature of

the problems and, indeed, of their solutions had changed dramatically.

Today's Grand Challenges are global and involve health, climate change and global warming, the environment, predicting and living with extreme seismic activities and tsunamis, disaster reduction and security and the spread and containment of infectious disease. These are fundamental issues that effect life on the planet, not just the profitability of the manufacturing industry. They demand profound understanding of exceedingly complex systems and they simply cannot be solved by teams constrained to a department, institution or organisation.

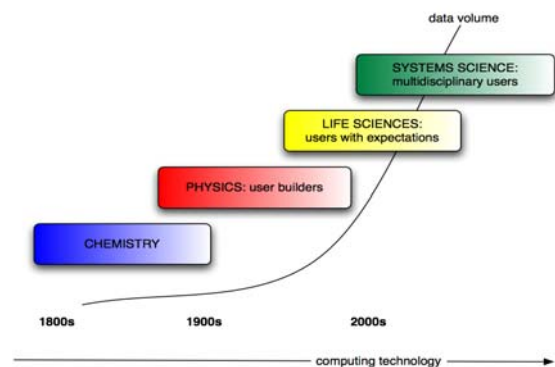


Figure 1: The change in science focus with technology.

Figure 1 illustrates the continuum of change in research with the influence of ICT and transition to meet new challenges over time.

2. System Science

Science has become a team sport and we have entered the era of “System Science” - the integration of diverse sources of knowledge about the constituent parts of a complex system with the goal of obtaining an understanding of the system's properties as a whole. System Science is characterised by global collaboration, shared resources, common goals and multi-disciplinary problem solving. Ian Foster, one of the recognised “fathers of the grid” provides the following commentary:

“System-level science integrates not only different disciplines but also, typically, software systems, data, computing resources, and people. System-level science is usually a team pursuit. Data comes from different sources, different groups develop component models, team members provide specialised expertise, and the often substantial computing and data resources required for success are themselves diverse and distributed. Thus, system-level science itself requires the creation of yet another sort of system that may combine large numbers of both physical and human components.”ⁱ

An example of System Science is a virtual organisation (VO) for the study of ecoinformatics: an online, collaborative and shared resource for managing ecosystem and environmental data and information products. The ecoinformatics VO becomes the focal point for research collaborators and policy participants: a common, virtual platform that provides a consolidated data and knowledge infrastructure to support research design and delivery, and greatly improved collaboration across research teams in a shared environment. It brings together, as a base set, data and methods from ecology, environmental science, land care, water management, climate studies, agriculture, animal husbandry, mathematics and computer science and advanced ICT.

Treating ecoinformatics as a System Science will lead to innovative methods for the integration, curation and management, mining, interrogation, visualisation and modelling of multi-scaled (local, regional, national and global) data within the shared platform, thus delivering the best available and most cost-effective research outputs to the natural resource and primary industry sectors.

The growing awareness that researchers must collaborate, that resources have to be shared, that information and communication technology has to be better integrated into the research environment and that these characteristics apply to all research activities not just science, has resulted in the genesis of eResearch.

3 eResearch in Australia

Dr Mike Sargent, Chairman of the Australian e-Research Coordinating Committee (eRCC), Australia's equivalent of Sir John Taylor, wrote in his report to the Government:

“On a global scale, the level of investment in ICT enablement of research in the United Kingdom, United States of America, Europe and Canada is beginning to pay dividends. In the United Kingdom, £250M has been invested in the e-Science programme over the past 5 years and the programme has drawn further significant investment from corporations such as IBM, Microsoft and Oracle. In the Asia Pacific region, the governments of Japan, Korea, Taiwan and China are strongly committed to finance their e-Research frameworks. Countries investing in e-Research capabilities are increasingly participating in collaborative research into the grand challenges.”ⁱⁱ

Australia has been slow to adopt eResearch, however, this tardiness has been an advantage in a country with merely twenty-one million people and fewer than eighty thousand researchersⁱⁱⁱ. Australia has been able to watch and learn from the development of eResearch activities in the US, Europe, UK, Japan, Korea, China, Singapore and Latin America and develop a philosophy of “adapt – adopt – and contribute” based on successes elsewhere.

In particular, Australia has modelled its eResearch development on that of the UK. A large number of delegations have explored every aspect of UK eScience over the last five years and several key emissaries, for instance, Sir John Taylor, Dr A. Hay, Professors A. Trefethan, C. Gobles, and D. DeRoure, have visited Australia to advise on eResearch policy and procedures.

VerSI, the Victorian eResearch Strategic Initiative, exists on a much smaller scale than the UK wide eScience Program. Nevertheless, the principles of awareness as the first step; working with the research community to establish trust; developing solutions for selected research groups that facilitate collaboration and tangible examples of the benefits of

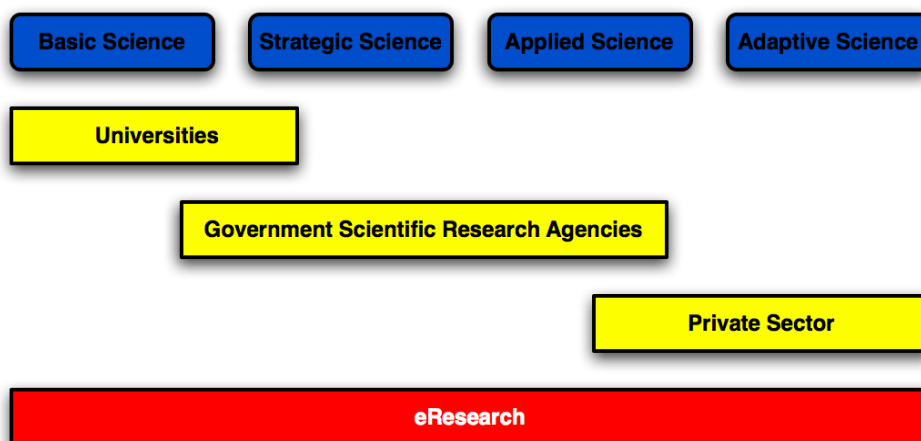


Figure 2: Science value chain

collaborative research and shared resources, and; providing key enabling infrastructure, are the same.

The Australian Department of Education, Science and Technology (DEST)^{iv} proclaims:

“The research sector worldwide is experiencing enormous change driven by advances in information and communications technology (ICT). Research is increasingly characterised by national and international multi-disciplinary collaboration and most OECD countries and APEC members are investing heavily in those capabilities and the associated coordinating mechanisms.

The term ‘e-Research’ encapsulates research activities that use a spectrum of advanced ICT capabilities and embraces new research methodologies emerging from increasing access to:

- *Broadband communications networks, research instruments and facilities, sensor networks and data repositories;*
- *Software and infrastructure services that enable secure connectivity and interoperability;*
- *Application tools that encompass discipline-specific tools and interaction tools.”*

Apart from the Australian Government’s insistence, there are several compelling reasons why eResearch is becoming the new paradigm. Robert Kelley (Carnegie Mellon University) claims that, in 1986, 75% of the knowledge one needed to do one’s job was stored in the mind. In 1997 this had fallen to 15% to 20%^v;

knowledge has become an on-line commodity. Howard Garner (Harvard University) amplified this message when he stated, *“Knowledge does not stop at my skin: it includes my computer and its databases and my network of associates”^{vi}*. eResearch provides the infrastructure and methodologies for accessing and reusing on-line knowledge.

Figure 2, above, is a schematic of the science value chain illustrating that eResearch, by virtue of the collaborative and shared dogma, provides a seamless communications path for ideas, data and knowledge from the fundamental research carried out (mostly) in universities through to industry.

Two years ago DEST commissioned the Australian eResearch Coordinating Committee to undertake a comprehensive review of eResearch and recommend how Australia, cognisant of experiences elsewhere, could coordinate a National eResearch initiative. One of the recommendations of the Committee was the establishment of an eResearch Centre consisting of a coordinating body and six state-based nodes to facilitate the transfer of eResearch methodologies to the research community.^{vii}

4 Challenges to eResearch in Australia

eResearch initiatives, such as VerSI, are not without their challenges; several key issues in the adoption of eResearch have yet to be addressed. These include the reward system for academics, which is not geared to adequately measure outputs other than papers.

Encouragement of open-source software, on-line publication and sharing of experimental data and derived results all test the ability of the research sector to measure real research outputs and reward researchers accordingly. This is particularly important as the research community is encouraged to move from competitive to collaborative endeavours.

Before researchers will adopt new methods there has to be a clear, unambiguous and demonstrable benefit. In chemistry, the reaction coordinate is an abstract one-dimensional coordinate system that represents progress along a reaction pathway (Figure 3). Reactants are mixed, energy is added (the activation energy) and the reaction proceeds to the formation of the wanted products. If insufficient energy (less than the activation energy) is added the reaction cannot proceed. By analogy, for a researcher to move to a new paradigm the “product” environment must be better than the starting point AND the “activation energy” (learning, re-training, disruption to research activity, costs of change etc) has to be very small. VeRSI’s role is to describe and demonstrate ways to adopt eResearch that lead to better research environments and require little or no energy input from the researchers. The hidden catch is in the subjective term “better”; in this context “better” is determined by the individual researcher.

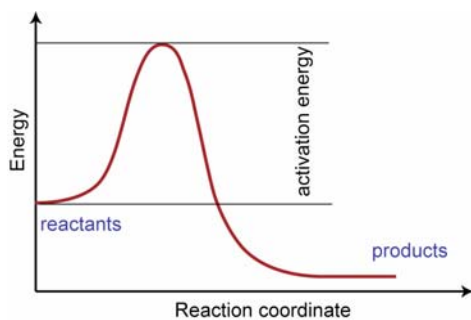


Figure 3: Chemical reaction pathway – an analogy for change

The most significant challenge, however, is the paucity of skilled ICT personnel, particularly those with multidisciplinary expertise, and the difficulties of recruiting staff who can bridge the void between research disciplines and technology developers. This was clearly identified by the eRCC who recommended the establishment of an extensive training program involving a five-year programme of one-year eResearch Honours Scholarships and a five-year programme of three-year eResearch Postgraduate Scholarships to support students

undertaking research training in eResearch capabilities at the PhD level. It was proposed that this be supplemented with the introduction, by institutions, of formal incentives, recognition and reward mechanisms for the skilled eResearch professionals who support and provide expertise to researchers, and to encourage academics to invest time and expertise to develop cross-disciplinary projects and courses to train a new generation of eResearchers.

The press (iTWire) report that:

“Salaries in the Information, Communication and Technology sector were up by 12% across the board for the six months to December 2006...”

and, *“...demand for application developers, data management professionals, and business analysts was continuing unabated with this trend expected to go on strongly in 2007.*

“The market has been, and remains, hot with demand outstripping supply,” Andy Cross, Ambition technology managing director said. “Salaries have jumped 12% and the technology recruitment market can’t bridge the demand gap. That means the war for talent will continue in 2007. This will be the key challenge in our sector for next year.”^{viii}

With this level of competition and the uncompetitive salaries and conditions offered by the university sector the recruitment challenge will continue to be the major gap between the ideals of eResearch and its realisation.

5 eResearch in Victoria

The Victorian State Government, through Multimedia Victoria, decided to meet these challenges and in 2006 announced funding for the Victorian eResearch Strategic Initiative (VeRSI) as part of its Healthy Futures: Life Science Statement^{ix}. The government funding will be supplemented by equal amounts of in-kind support from the partners. The clear intention is that VeRSI become the Victorian node of the national eResearch Centre.

The VeRSI Program is an unincorporated joint venture. The initial members of the Consortium are Melbourne University and Monash University along with the Victorian Government. VeRSI is based on collaborative and inclusive model and will openly involve research groups, skills and services from other universities, research organisations, government departments and service providers such as the

Australian Synchrotron, Victorian Partnership for Advanced Computing (VPAC) and Victorian Education and Research Network (VERNet).

The funded activity will proceed from October 2006 to September 2010. During this period the VeRSI Program will undertake five complementary projects to provide support and services to researchers, undertake an extensive outreach and awareness raising exercise, deliver a coordinated program of skills development and assist in the coordination of research development and deployment.

The five projects are broken into three categories: enabling projects, a demonstration project and capability projects.

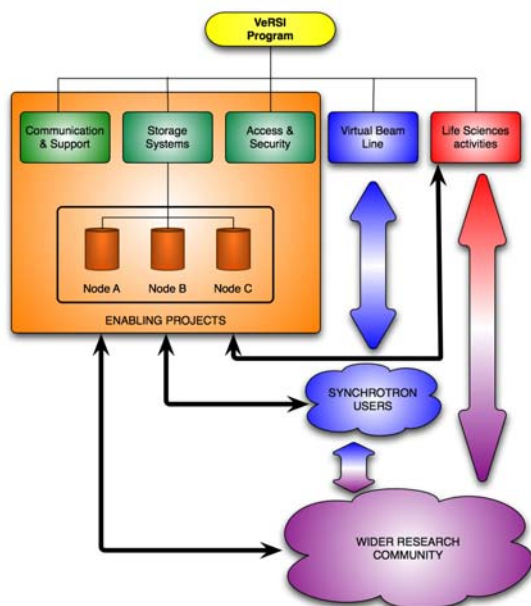


Figure 4: Relationships between the VeRSI projects

5.1 Enabling Projects

The enabling projects will create awareness and establish essential infrastructure such as an access and security framework and distributed, federated storage facilities. There are three enabling projects:

- **Communications & Support:** Communications are a critical component in the development of an eResearch framework. Informing researchers, explaining the benefits of eResearch methods and assisting them to assess their requirements and general engagement are all key, on-going activities that form the Awareness, Outreach, Communications and Support project.

- **Security & Access:** Security and access includes the means to identify and authenticate users, to provide them with access to shared resources and to provide assurances that information stored in shared repositories is safe and secure. The challenge is to balance the security and access needs of the community of interest with those of the host institution(s). VeRSI are developing a standards-based security and access policy that is compatible with existing systems and implementing the supporting ICT platform so that authenticated access is possible from anywhere.

- **Storage Systems:** Data storage is a high demand commodity and most university systems are over taxed and designed specifically to meet the immediate needs of the campus. This leaves little scope for experimentation with storage standards and protocols. The UK eScience program successfully addressed this issue by providing limited, production quality resources, particularly data storage and management, as a foundation on which to build trust and deliver services with minimal impact on University security systems and firewalls. VeRSI will follow this example.

5.2 Demonstration Project

The demonstration project is the design and construction of a prototype Virtual Beam Line (VBL) to demonstrate remote working with the High-throughput Protein Crystallography beamline at the Australian Synchrotron. The VBL will allow for collaborative working as part of grid-enabling the Australian Synchrotron and synchrotron user community. The VBL will provide a multimedia nexus through which users can collaborate using voice, video and shared applications, undertake occupational health and safety, radiation and beamline training, use a variety of tools to transfer data to external storage and computing resources and remotely monitor and mentor their experiments.

The VBL design will be modular, easily replicated and meet the need of the synchrotron community. The design will be promoted and freely available to interested parties at other beamlines.

5.3 Capability Project

The capability project is a series of exemplar use-cases and applications for researchers in the Life Sciences, capitalising on Australia's existing research strengths. These projects will deliver production quality tools to the specific research groups and serve as tangible examples to the research community of how advanced

ICT that is responsive to researchers' needs can enhance the research environment. There are eight use-cases:

- **Genomic dataset mining:** The function of genes is dependent on their structure but, due to the size of genomic data sets, mining of the gene structure is an arduous task. Further insights are developing from the merging of publicly available data with proprietary information. Innovative methods for collection, curation, visualisation, integration and mining of these data will deliver a major advantage to the Victorian biotechnology and biomedical research sectors.

VeRSI is supporting this research by designing the general architecture of software modules for data mining, assisting the implementation of novel data mining techniques and building a user interface to the data sets.

- **Neurosciences & biomedical imaging:** The Neuroimaging Laboratory at the Howard Florey Institute is using human and animal brain imaging techniques to investigate clinical and basic neuroscience research questions. These include how various functions change with variations in hormone levels, and whether neurobiological changes can be observed prior to the onset of physical symptoms in sufferers of diseases such as Huntington's or Alzheimer's disease.

VeRSI is supplying crucial integration and networking capabilities, in addition to building a metadata explorer model, a metadata editor for neuroimaging data sets, a user interface to the data and a data authentication model.

- **Australian Mouse Brain Map:** The Mouse Brain Map Consortium is building a national atlas of brain data, from MRI, histology and immuno-histochemistry, for comparison between diseased mouse brains and control brains, studying changes in brain anatomy and function.

VeRSI is building a repository of mouse brain map images and information from institutes within the consortium, designing the general mouse brain map database architecture along with dataflow, specimen tracking and security models and building an interface to the repository for upload, management, searching and download of data.

- **Uro-oncology Informatics Grid:** Around fifty percent of Australian men experience some type of prostate problem during their lifetime. Prostate cancer is one of the most common

forms of cancer in men. The Centre for Urological Research aims to provide better diagnosis and treatments for prostate cancer and benign prostate disease. A biorepository of prostate, kidney and bladder specimens will allow researchers to share resources with existing national and international repositories.

VeRSI is designing and developing an informatics grid, ensuring the informatics grid is interoperable between research centres and national and international tissue banks, and organising lab data, urology data, and pathology data.

- **Ambulatory motion studies:** New technology allows the easy collection and examination of numerous gait parameters, over many trials, for large groups of people within a laboratory or field setting. The data provides clinically valuable information about when gait changes occur, leading to a improved understanding falls and serious injury in older adults.

VeRSI is improving the management, distribution and access of gait motion study data by designing and building the data architecture for storage of gait study data and building a security model for access to data sets.

- **Workflows and laboratory automation for metabolomics:** The fundamental challenge of the post-genomic era is to utilise the information generated by genome sequencing projects in conjunction with high-throughput profiling technologies to understand cellular functions on the molecular level in a comprehensive and integrated manner. Metabolomics is an emerging tool increasingly used to identify new protein functions and to model the whole cell metabolism.

VeRSI is developing workflows and data management models to automate metabolomics research at the Bio21 Molecular Science and Biotechnology Institute, designing a data storage model for metabolomics data to capture the entire data flow and designing and implementing a web based user interface to enable data management and provide access to shared data.

- **Distributed radiotherapy system:** Computation for radiotherapy patient treatment planning requires high performance computing (HPC) resources, but it is preferable for patients to be treated as close to home as possible. Distributed treatment planning allows for planning to occur at HPC resources in

metropolitan areas after which it can be distributed to regional clinics.

This use-case will provide a connection between Peter MacCallum Cancer Centre and satellite centres, via VERNET (Victorian Education and Research Network), and develop an authenticated portal interface to treatment planning, imaging and clinical datasets.

- **Laboratory data management:** Australian scientists need to build a critical mass and develop workflow technologies and practices that link geographically dispersed groups. This will enhance collaboration in areas of niche strength such as agrifoods biotechnology, stem cell research, synchrotron-based research and advanced clinical trials. It can be achieved by combining internet-based collaborative environments and grid computing, collectively known as eScience, which in turn will underpin the enhanced data handling and mining methods.

This use-case will develop workflow technologies and practices that link geographically dispersed groups and enhance collaboration in areas of niche strength such as agrifoods biotechnology, stem cell research, synchrotron-based research and advanced clinical trials

5.4 VeRSI Outcomes

VeRSI activities will provide support and services to researchers, undertake an extensive outreach and awareness raising exercise, deliver a coordinated program of skills development and assist in the coordination of research development and deployment.

The outputs from the project groups will be knowledge and expertise, designs and technology solutions and advanced open source software. All of these commodities have value that will enhance the quality of research, bring timelier research outcomes, catalyse international collaborations, and provide opportunities to Victorian industry.

VeRSI will deliver research leadership by harnessing enabling technology in a way that delivers a productive, collaborative research environment and by developing key examples of how eResearch can enhance research outcomes. This approach will catalyse the widespread adoption and uptake of eResearch and deliver the promise of faster and more exhaustive research activities leading to improved commercialisation opportunities and an elevation in the status of the State's academic

and research institutions. It will add value to the Australian Synchrotron, improve collaboration in the Life Sciences and reinforce Victoria's position as a knowledge-based economy.

6 Conclusion

The adoption of eResearch is a cultural change process which, paradoxically, will have succeeded when the "e" is no longer necessary and the "research method" encompasses all the tenets of shared, collaborative working enabled by advanced information and communication technologies. But this will take time. We can learn the sociological, physiological and technological lessons of the UK, Europe and the US all of whom lead Australia in the adoption of eResearch but this will have little effect on the speed with which the culture of research changes. The "e" will be part of the promise for a time to come.

ⁱ I. Foster, C. Kesselman, " Scaling System-Level Science: Scientific Exploration and IT Implications." *Computer Magazine*, vol. 39, no. 11, 2006, pp. 31-39.

ⁱⁱ eResearch Coordinating Committee, *Final Report of the e-Research Coordinating Committee*,⁵ DEST, http://www.dest.gov.au/sectors/research_sector/publications_resources/profiles/e_research_strat_imp_framework.htm, 2004)

ⁱⁱⁱ *Research and Experimental Development, All Sector Summary, 8112.0*, Australian Bureau of Statistics,

^{iv}

http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/e_research_consult/default.htm

^v R.E. Kelley, *How to Be a Star at Work: 9 Breakthrough Strategies You Need to Succeed*, Three Rivers Press 1998,p77.

^{vi} H. Gardner, *Intelligence: Multiple Perspective*, Wadsworth Publishing 1995

^{vii}

http://www.dest.gov.au/sectors/research_sector/publications_resources/profiles/e_research_strat_imp_framework.htm

^{viii}

<http://www.itwire.com.au/content/view/7503/50/>

^{ix}

http://www.business.vic.gov.au/BUSVIC/STAN/DARD//pc=PC_61353.html