Mind the Gap

Paul Davis

Abstract

VeRSI – the Victorian eResearch Strategic Initiative was funded in 2006 by the Victorian Government. This presentation will describe how the VeRSI program is providing a cohesive and coordinated approach to accelerating the uptake of eResearch by Victorian researchers.

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. The paper will describe the five complimentary 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. The paper will highlight the need to bridge the skills and expertise ‘gap’ between research disciplines and technology implementations.

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 the efficiency of 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.

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.”*

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 & 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.

The Australian Department of Education, Science and Technology (DEST) website 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 Government's insistence there are several compelling reasons why eResearch is becoming the new paradigm. Robert Kelley (Carnegie-Mellon) 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%; knowledge has become an on-line commodity. Howard Garner (Harvard) amplified this

message when he authored a white paper entitled, *"Knowledge does not stop at my skin: it includes my computer and its databases and my network of associates"*. eResearch provides the infrastructure and methodologies for accessing and reusing on-line knowledge.

Figure 1 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.

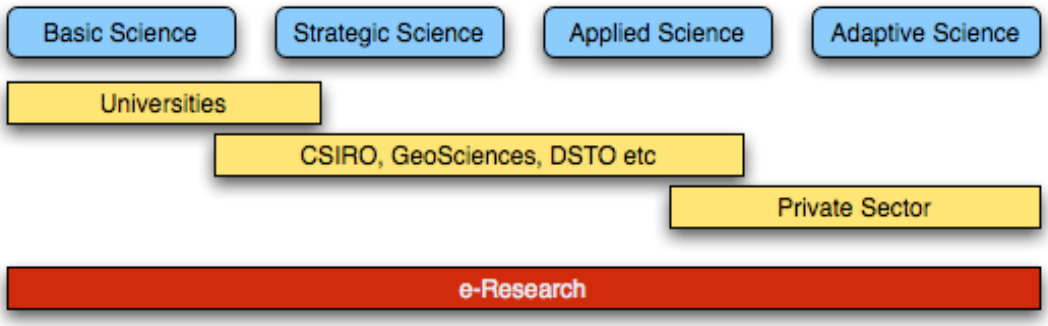

Figure 1 Science value chain

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.

The Victorian government, through Multimedia Victoria, had been thinking along similar lines and in 2006 announced $4.75M of funding for the first phase of the Victorian eResearch Strategic Initiative (VeRSI) as part of its Life Science Statement. The government funding will be supplemented by approximately $5M of in-kind from the partners and participants. 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, however, 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 complimentary projects:

enabling projects, to create awareness and establish essential infrastructure such as an access and security framework and distributed, federated storage facilities;

a development project, a Virtual Beam Line for collaborative working as part of grid-enabling the Australian Synchrotron and synchrotron user-community, and;

capability projects, and series of exemplar use-cases and applications for researchers in the Life Sciences. 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 aware of researchers needs can enhance the research environment.

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.

Initiatives such as VeRSI are not without their challenges and 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 2). 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.

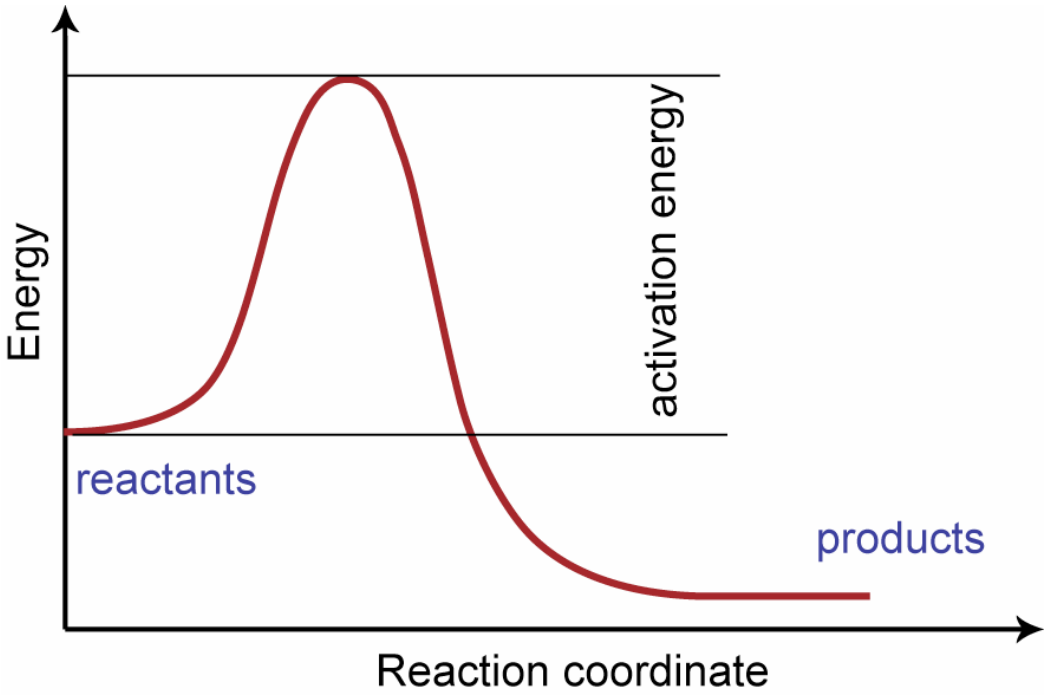

Figure 2 Chemical Reaction pathway – an analogy for change

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 program involving a five-year programme of one-year eResearch Honours Scholarships, a five-year programme of three-year eResearch Postgraduate Scholarships to support students undertaking research training in eResearch capabilities at the PhD level and 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 http://www.itwire.com.au/content/view/7503/50/) 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.""*

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.

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.