Professor Richard O. Sinnott
TECHNICAL DIRECTOR NATIONAL E-SCIENCE CENTRE, UNIVERSITY OF GLASGOW & INCUMBENT E-RESEARCH DIRECTOR, UNIVERSITY OF MELBOURNE

ARDR OPINION

Research in the digital age

The digital age is truly upon us and this is directly influencing the way we communicate, socialise, shop, travel, educate and indeed undertake research. Whilst offering numerous opportunities, this technical revolution has numerous challenges that must be overcome. The size of the internet as a communications vehicle is growing beyond all expectations, with zettabytes (10^{21} bytes) of data expected to be reached in 2010.[1]

This challenge is exacerbated by the way in which research is undertaken and automated. From high throughput post-genomic experiments generating terabytes of data per experiment as genomes can now be sequenced at hitherto unprecedented speed; through searching for 'needle in the haystack' particles by analysing petabytes of data (10^{15} bytes) coming from experiments such as the Large Hadron Collider in CERN; to projects in the humanities and the social sciences where heterogeneous and voluminous data sets are growing and require storage and long term preservation to both underpin research and societal demands.

There are indeed few research areas where the growth of digital data and the capacity for processing of this data is not a common problem that is now actively hindering research progress. Technologies and solutions do however exist to support many of these data challenges. One only has to look to internet search engines such as Google to see how rapid searching of data can be supported through building of indexes of publicly accessible web pages. However, whilst useful and powerful, these generic tools often do not meet the needs of the wider research community. Thus when data is kept in targeted databases or in secure settings (e.g. clinical data environments as typified by the Parkville Precinct and centers such as Bio21 and the Peter MacCallum Cancer Research Centre) then use of general purpose solutions like Google rapidly breaks down.

Yet computer capacity and provision of peta-flop (10^{15} floating point instructions per second) high performance computing (HPC) systems are continuing to withstand Moores Law on the scaling of computing power and performance. The Victoria Life Sciences Computation Initiative is testament to this!

There are indeed few research areas where the growth of digital data and the capacity for processing of this data is not a common problem that is now actively hindering research progress.

The marriage of research-driven challenges which embody large scale data volumes and the need for access to HPC resources has driven many efforts in the area of e-Science. Often these 'big science' models and supporting IT solutions (often realised by Grids), share many desirable features common to many research areas. These include the need for seamless access to heterogeneous, distributed data sets from autonomous providers; the use of computer facilities for simulations/data processing; and support for single sign-on, where a single authentication/authorization can be used to access resources distributed across multiple sites. Hence e-Science has effectively morphed into e-Research.

E-Science is not solely big-science however. At the heart of the problem facing many researchers in many research areas is collaboration and co-ordination. How can researchers find data resources that are not publicly accessible? How can these researchers and/or associated data providers share resources without necessarily giving away their data, their intellectual property or their research results? How can they run large scale simulations at multiple scales, for example from the cell to the organism, from the atomic to the macroscopic, without themselves having to become HPC experts? How do they manage their data long-term?

E-Science is not solely big-science however. At the heart of the problem facing many researchers in many research areas is collaboration and co-ordination.

Provisioning of e-Infrastructures for such scenarios is a cornerstone of e-Research. Often this is achieved through the definition and support for virtual organisations where users and resource providers agree on the framework for collaboration and technical solutions that can be used to orchestrate and enforce this. Fundamental to the success of these virtual organisations is capturing the domain knowledge and requirements of all participants and concerns of associated stakeholders.

This is especially so given that virtual organisations can be (and indeed often are!) inter-organisational and inter-disciplinary.

Taking the example of e-Health and on-going virtual organisations at the National e-Science Centre (NeSC) at the University of Glasgow: a multitude of domain experts including clinicians, biologists, pharmacists, physicists, chemists, social scientists, biostatisticians, epidemiologists, often need to be involved and input their expertise into a given research collaboration and have timely access to the right data at the right time. Such models are typified by the need for rigorous fine-grained security solutions that cross inter-organisational boundaries where all protagonists are aware of their roles and involvement in the collaboration as a whole.

These models have been applied in multiple forms in a myriad of research areas at NeSC – from commercially sensitive and industrially oriented nanoCMOS electronics domain; the social simulation and social data management domain; clinical trials and epidemiological studies in the area of rare diseases, brain trauma, genetic causes of hypertension, adrenal cancers; prediction of Parkinson's disease; geospatial information systems through to the arts and humanities and the language and literature domain. The scope and potential for e-Research and establishment of multiple new virtual organizations is immense.

From past experience, successful collaboration demands e-Infrastructure developers and providers embed themselves in application domains and actually listen to the demands of the researchers themselves (as opposed to simply pushing technology). The success of NeSC at Glasgow throughout the full course of the £250m+ UK e-Science cross research council core programme is testament to this.

I fully expect to continue with an e-Research engagement and uptake formula that works in, and for, Melbourne and Victoria!

[1]Digital information will grow to 1.2 zettabytes this year: IDC study, http://www.kurzweilai.net/news/frame.html?main=news_single.html?id%3D12119 , May 2010.