The International Journal of Digital Curation
Issue 2, Volume 2 | 2007

Report from the UK e-Science All Hands Meeting, September 10-13, 2007, Nottingham, UK

Colin Neilson,
UKOLN, University of Bath

November 2007

Summary

This is a report from the UK e-Science All Hands Meeting held at the East Midlands Conference Centre, Nottingham on September 10-13, 2007. This event was the 5th annual All Hands Meeting. The proceedings of the meeting are available from the UK e-Science 2007 All Hands Meeting Website.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

The All Hands Meeting (UK e-Science AHM, 2007) boasted a delegates list of more than five hundred people from an impressive range of organisations. The final programme included 51 reviewed papers, 18 workshop papers and 30 poster papers. The effort made in managing the quality of submissions can be ascertained from the list of 78 other reviewers listed in the proceedings as well as the 43 members of the AHM07 programme committee.

The meeting was structured around workshops, main conference keynote addresses, parallel sessions at which papers were delivered, tutorials and poster sessions. Given the complexity of parallel presentations, and despite the compact and functional setting of the East Midlands Conference Centre, it would not be possible (or indeed wise) for one person to attend or summarise all that happened.

This report reflects a subjective view of what was interesting for someone who is not an e-Science practitioner. The first impression was that the sheer number of practitioners dominated the event; it felt like a big meeting. To some extent the impression supported the claim for e-Science as an important discipline with a mature programme of work, as made by **Professor Malcolm Atkinson** in his welcome and opening keynote address on the future of e-Science.

What Is Meant by "e-Science"?

The UK e-Science Programme began in 2001 as a five-year initiative with an e-Science Core Programme managed by the EPSRC (Engineering and Physical Sciences Research Council) aiming to serve communities of all the UK research councils and the e-Science research programmes run by each research council. One aim was to support the development of generic technologies, such as middleware software, allowing different resources to work in a seamless way across networks. The vision was very much an "on demand" service and data access through a computing "grid" environment. The aim being to tie high availability and performance of computation, communications networks and large-scale data together as a service to support a wide range of academic, industrial and commercial research. For a considered account of the definition and development of e-Science, see Jankowski (2007) in the special issue of *Journal of Computer-Mediated Communication* devoted to e-Science.

The AHM07 Workshop 1 on OMII-UK (the Open Middleware Infrastructure Institute UK) illustrated the development of such software tools and services and the maturity in methods of software engineering adopted, and how these could be sustained including the OMII-UK Open Forum for users.

As **Professor John Wood** highlighted in his keynote address, some aspects of the UK e-Infrastructure are now in place and the UK research councils may regard these aspects as work done. This would mean the future funding for e-Science could be integrated with other initiatives. The challenge is how one appropriately makes that infrastructure a part of researchers' everyday work.

To some extent the variety of possible uses of e-Science across a full range of disciplinary approaches makes e-Science difficult to understand. AHM07 Workshop 2

on "Building Usable Systems for the Global Environment" provided examples in different domains (e.g. Astronomy, Healthcare research) of deployment of systems for operational and scientific use. How have projects shared techniques and solutions across domains making sure that systems are "usable" to their communities? How does international collaboration function in relation to inter-operability of systems of working? Many of these issues are only easily discussed and understood in the context of examples from particular projects. It is not clear that there is any general cross-domain solution to "sharing" data as a simple issue of technological capability. The data sharing exercise may only make sense in the context of the particular research field where it can be shown to have a positive effect on research outcomes. Of course there is a larger idea of "Data-driven" research impelling interest and support for e-Science (see **Murray-Rust** (2007) for one scientist's exposition of this idea).

Digital Curation

Often when "data sharing" or "data re-use" is discussed, what is actually implied is a common purpose based on shared understanding, shared meanings and perhaps even common collection and systems of coding, as well as an understood ethical or legal framework within which rule-based compliance operates. Even the more technical aspects of data curation, such as choice of metadata to be collected, require a supporting context and understanding of the research culture of each domain. The workshop 6, "We have to talk about metadata" provided a number of examples of development in specific domains. However there was underlying uncertainty expressed about what metadata was exactly and how it is distinct from data and also a number of concerns expressed about the burden of work implied in any extensive collection of metadata. The workshop 4 on "Issues in Ontology Development and Use" provided a review of approaches to capturing, and representing in a computer-based form, the semantic information required to allow researchers to take increased value from digital working.

Many of the issues around data sharing, and digital curation in general, may need to be socially defined. The presentation on "A Socio-technical Perspective on Ontology Development in HealthGrids" by **Jenny Ure** (Ure, 2007), provided a valuable account of experience in a health and clinical setting. The technical infrastructure of e-Science is necessary but not sufficient in meeting the needs of researchers in getting full value from digital working. The "human infrastructure" behind research work in each domain needs to be better understood.

The SUPER project (Newhouse, 2007) identified many issues to be addressed in the future development of e-Science. These included standard policies and tools for data curation across Research Councils UK-funded research. Additional tools were needed to autogenerate metadata including the provenance of data.

I found the easiest route at AHM07 to understanding the benefits of e-Science was through the reports based in the medical domain. The presentations in the session on Medical Applications were very helpful. The CARMEN Project (Watson, 2007), supporting communities in neuroscience, reported on automation of a scientific process replacing significant amounts of manual effort. The analysis services made available allowed access to the data for other research. This demonstrated data re-use based on a repository service. It was intriguing to speculate how this would transfer to other domains. In any field reaching for mature development there is an expectation

that lessons can be shown to have been learnt. The PsyGrid Project (Ainsworth, 2007), working on clinical e-Science in the mental health domain, took a lessons-learned, evaluation approach making the case for user-centred design.

For me, the single most impressive presentation at AHM07 was the keynote address by **Professor Ander Ynnerman** on "Medical Visualization Beyond 2D Images". The idea of volume rendering, display of image banks and virtual reality experience does not convey the excitement of seeing a presentation on Forensic Virtual Autopsies by Direct Volume Rendering at this meeting (for background on this research see Ljung (2006)). The Virtual Autopsy used Multi Detector Computer Tomography data performing rapid scans (under 30 seconds) of cadavers leading to upwards of 8000 images and a data set of several gigabytes for each cadaver. The techniques used for direct volume rendering produced a detailed layered visualisation of the bodies allowing investigation of causes of death. The Virtual Autopsy was able to complement any requirement for physical autopsy and in some sensitive cases (e.g. investigation of the death of a baby) replaced the need for physical autopsy.

Technically one of the interesting features was the use of commodity-based graphics technology. The ability to re-examine the virtual autopsy at any point, including remote consultation of experts, was an unusually direct demonstration of the benefits of digital working and lent an entirely fuller meaning to the concept of data re-use and visualisation in support of expert judgment.

How Far Can e-Science Progress?

My main impression from the All Hands Meetings was of the wide extent, high level and variety of work undertaken across many domains. It seemed to me to be a very ambitious research and service development agenda. I sometimes found it difficult to distinguish "proof-of-concept" projects from those genuinely intended to deliver ongoing services to their specialist subject domains.

What is being proposed, and in some cases demonstrated, is automated life cycle management of research including attribution and provenance of the full set of research outputs. The benefits will include the provision of wider access to raw and processed data permitting analysis and re-use, thereby increasing the value to be gleaned in exploiting research.

My own view is that the development in different domains will continue to be uneven. I think the systems of management in many fields of research are not yet sufficiently mature to ensure cross-domain research capabilities. This includes quality management systems in working with data (including instrument-gathered data), information security management systems, and systems governing digital curation practice. Some approach to capability maturity assessment for digital curation will be required to foster the fullest development of automation and exploitation of benefits.

However the point of meetings such as AHM07 is to highlight how far e-Science has progressed over such a short period. Some of the projects covered are definitely a beacon for progress. I would recommend attendance at the future AHM08 meeting for researchers outside e-Science as a valuable and effective way to gauge the state of the art.

References

Ainsworth, J. (2007). The challenges of clinical e-Science: Lessons learned from psygrid. In *Proceedings of the UK e-Science All Hands Meeting 2007, Nottingham, UK, September 10-13, 2007*. Retrieved October 31, 2007 from http://www.allhands.org.uk/2007/proceedings/papers/880.pdf

Jankowski, N. W. (2007). Exploring e-science: An introduction. *Journal of Computer-Mediated Communication,* 12*(2), article 10*. Retrieved October 31, 2007 from http://jcmc.indiana.edu/vol12/issue2/jankowski.html

Ljung, P., et al. (2006, September/October) Full body virtual autopsies using a state-of-the-art volume rendering pipeline. *IEEE Transactions on Visualization and Computer Graphics, Vol. 12, No. 5*. Retrieved October 31, 2007 from http://ieeexplore.ieee.org/iel5/2945/4015412/04015441.pdf?tp=&isnumber=&arnumber=4015441

Murray-Rust, P. (2007). Data-driven science –A scientist's view. In *NSF/JISC Repositories Workshop, April 10, 2007*. Retrieved October 31, 2007 from http://www.sis.pitt.edu/~repwkshop/papers/murray.html

Newhouse, S. (2007). Study of User Priorities for e-Infrastructure for e-Research (SUPER). In *Proceedings of the UK e-Science All Hands Meeting 2007, Nottingham, UK, 10th-13th September 2007*. Retrieved October 31, 2007 from http://www.allhands.org.uk/2007/proceedings/papers/784.pdf

UK e-Science 2007 All Hands Meeting Website (2007). *Proceedings of the UK e-Science All Hands Meeting 2007, September 10-13, 2007*, Nottingham UK. Retrieved October 31, 2007 from http://www.allhands.org.uk/2007/proceedings/

Ure, J. (2007). A socio-technical perspective on Ontology development in healthgrids. In *Proceedings of the UK e-Science All Hands Meeting 2007*, Nottingham, UK, September 10-13, 2007. Retrieved October 31, 2007 from http://www.allhands.org.uk/2007/proceedings/papers/873.pdf

Watson, P. (2007). The CARMEN neuroscience server. In *Proceedings of the UK e-Science All Hands Meeting 2007*, Nottingham, UK, September 10-13, 2007. Retrieved October 31, 2007 from http://www.allhands.org.uk/2007/proceedings/papers/785.pdf