



Birds of a Feather (BoF) Session on “e-Infrastructure: tool for the elite or tool for everybody?”

**held on the 8th September 2008 at the
All Hands Meeting 2008, University of Edinburgh**

Alex Hardisty, Cardiff University
Neil Chue Hong, University of Edinburgh
Rob Procter, University of Manchester
Lorna Hughes, King's College London
Alex Voss, University of Manchester
Mike Fraser, University of Oxford

05-Jan-2009

Abstract:

A Birds of a Feather (BoF) meeting brought together more than 70 interested participants of the e Science All Hands Meeting 2008 to consider questions and issues arising from the proposition: “e-Infrastructure: tool for the elite or tool for everybody?” Contrasting position statements from two practitioners in the field set the scene for multiple break-out groups to debate several questions derived from the proposition, concerning the nature, use and evolution of e-Infrastructure. Results showed considerable commonality of thought among the break-out groups, with 7 broad categories of response emerging. These categories were: i) There is no single common e-Infrastructure; ii) Ease of use is the initial barrier; iii) Dealing with complexity is complex; iv) Trust is important; v) Open development is necessary; vi) Give credit for digital creation; and vii) Attitudes must be changed. Within each of the categories a number of detailed points were collected. The results accord well with findings from interviews conducted by the ENGAGE project; namely that people will tend to prioritise ease of use, support and continued development over a complete feature set. This requires a sustainable community around the software and trust by the users in the e-Infrastructure providers (and vice-versa).

UK e-Science Technical Report Series

Report UKeS-2009-01

Available from http://www.nesc.ac.uk/technical_papers/UKeS-2009-01.pdf

Copyright © 2009 The University of Edinburgh. All rights reserved.

1 Introduction

A Birds of a Feather (BoF) meeting brought together more than 70 interested participants of the e-Science All Hands Meeting (AHM) 2008 to consider issues arising from the question:

“e-Infrastructure: tool for the elite or tool for everybody?”

The purpose of the BoF was to stimulate the AHM community (research scientists, computational scientists, computer scientists and software engineers) to actively discuss some of the key questions emerging from the JISC Community Engagement projects (e-Uptake¹, e-IUS², and ENGAGE³)⁴ and to consider the implications of these questions on community engagement with research computing and e-Infrastructure in the UK, and the sustainability of e-Infrastructure and tools. As a secondary objective, the BoF was intended to encourage greater community knowledge and sharing, and to lead to better engagement between researchers, providers, developers, funders and policy makers.

Alex Hardisty, (Cardiff University, Welsh e-Science Centre and Chair of the JISC Steering Committee for the Community Engagement projects) introduced the context of the BoF and the topics for discussion.

2 Topics for Discussion

Five main topics were suggested for discussion by the BoF organisers:

1. What do you think e-Infrastructure is and what should it be?

For example, is it a tool of use only for tackling the ‘grand challenges’ in research or could it (and should it) be useful for all kinds of research problem?

This question is about teasing out what e-Infrastructure is for, what it comprises, including the technical, social and organisational arrangements, and what this implies for future strategic development goals and priorities.

2. To what extent is the use of e-Infrastructure essential to your research?

Does it help you to tackle what you consider to be grand challenges in your research area? If it is optional but valuable then how do we motivate people to invest effort in adoption?

3. How would you use e-Infrastructure in the future?

What is the vision you have for your research in the future? If you are using e-Infrastructure today, how would you extend this use or make it more valuable? If you are not using e-Infrastructure, what would convince you to start using it?

¹ <http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/euptake.aspx>

² <http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/eius.aspx>

³ <http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/engage.aspx>

⁴ Part of the JISC e-Infrastructure programme, the Community Engagement projects are running from April 2007 to June 2009. Their results are being brought together in the [e-Infrastructure Community Engagement Portal](http://www.engage.ac.uk/) (<http://www.engage.ac.uk/>).

4. Do researchers need a clearly defined ICT environment and tool suite, or can usage be opportunistic, picking up on functionality that becomes available using light-weight “glue” and pragmatic organisational arrangements?

Should we build cathedrals or bazaars? i.e., should we wait for standards to build perfect systems or take what we have, get on with it and re-factor when needed? How does this affect reproducibility of results?

5. What would be needed to truly embed the use of e-Infrastructure in your work across the whole research lifecycle?

How should we measure effectiveness of the use of e-Infrastructures? What are the inspiring use cases? What are the drivers, both rewards and punishments, which would lead to the increased uptake of e-Infrastructure?

3 Position Statement Presentations

To set the scene, two “position statement” presentations were made by Jean-Claude Bradley (Associate Professor of Chemistry at Drexel University, Philadelphia) and Shantenu Jha (Senior Research Scientist at the Center for Computation and Technology and an Assistant Research Professor in the Computer Science Department at Louisiana State University).

Each speaker was asked to introduce their area of research interest and to explain how e-Infrastructure (by whichever definition they chose to use) plays a part in their work. They were asked to give their views on the priorities for investment in e-Infrastructure in order to create the biggest impact.

Jean-Claude Bradley introduced himself as a chemist. He stated his view that for some researchers and/or some projects the contents of lab notebooks should be available in real-time to anyone that wants to scrutinise them. There shouldn't be any “insider information” – information that is unavailable to others and which therefore requires others to trust what a scientist is claiming. Bradley calls this “open notebook science”,⁵ and in his presentation gave the example of the UsefulChem project⁶ to show how it can be accomplished using readily available Web 2.0 tools. These tools include: wikis, blogs, collaborative authoring of documents, slideshow and video sharing, social networking, online journal publishing, etc. without the need to worry about “software and servers”. Bradley noted that he is and wants to be a chemist, not a computer scientist! His view is that e-Infrastructure represents a great opportunity to make more use of the techniques of open notebook science. For some researchers and some projects there are advantages to doing science under open notebook science conditions. The main advantages are that it is a quicker way to share the science and to find potential collaborators. However, Bradley notes, it is not appropriate when there are novel intellectual properties to be protected.

Shantenu Jha introduced himself as a “card carrying computational scientist” with interests spanning multiple application classes. He stated his view that distributed applications of all kinds are still hard to do and that they are going to become even harder as a consequence of how computing infrastructures are evolving. He noted that, across the range of application classes, most difficulties arise because most applications can be either explicitly or implicitly distributed and that the approach taken in their design is often related to the extent to which the designers have control over the resources they choose to use. He suggested that, in order to be useful, the architecture of

⁵ http://en.wikipedia.org/wiki/Open_Notebook_Science

⁶ <http://usefulchem.wikispaces.com/All+Reactions>, <http://usefulchem.blogspot.com/>

e-Infrastructure needs to be designed by thinking from the applications perspective (rather than from the resource perspective). This means thinking about the interoperability of applications. It means identifying the right patterns and abstractions that could help to make the writing of distributed applications easier. In Jha's view, e-Infrastructure is anything that can be used to "compute" distributed applications, and it should consist of everything that is needed to enable that.

4 Break-Out Groups

Using the questions outlined above, the BoF participants divided into 13 mini-breakout groups, facilitated by the session leaders, to discuss and delve deeper into people's perceptions, criticisms and desires pertaining to e-Infrastructure. The make-up of each group was mixed, comprising users of different levels (expert, not so expert, etc.), developers and funders. Most individuals were UK based but the groups included some international representation.

5 Feedback from Break-Out Groups

5.1 Overall

Overall, from each of the 13 groups, there was significant commonality of thought. Ease of use, catering for a range of users, the need for durability of the e-Infrastructure, and the need for ubiquity of e-Infrastructure were common points made by many participants.

Similarly, many questioned whether it is possible to establish a single common infrastructure catering for all needs. However, many present felt that there would be many capabilities / functionalities of an e-Infrastructure that would generally be useful to a majority of the potential user base.

5.2 *What do you think e-Infrastructure is and what should it be?*

In response to the questions: "What do you think e-Infrastructure is and what should it be?", the following comments were generally made:

e-Infrastructure is:

- An application level stack giving access to grid hardware
- A tool layer that allows easy access to resources
- Portals providing an assembly point for using resources
- An information infrastructure giving details about what resources are available and what they can be used for
- The coordination of distributed tools enabling new ways of working
- Mostly the software bits above the cables but the rest (including non-technical bits) is important
- e-Infrastructure is the thing which allows researchers to use the resources available to them to carry out their research

e-Infrastructure should be:

- A secure way of working for projects using resources
- Of common use to some extent, as a single application for a single user is not infrastructure
- Providing a broad range of opportunities for as many users as possible
- Able to tell you what is available and what is appropriate for your use
- Mostly hidden from users

- But not a brick wall barrier for expert users
- Usable!

5.3 Making e-Infrastructure Mainstream

Taking the remainder of the discussion questions (2 above) under the general notion of “Making e-Infrastructure Mainstream”, we have grouped responses from the break-out groups into seven broad categories:

1. There is no single common e-Infrastructure
2. Ease of use is the initial barrier
3. Dealing with complexity is complex
4. Trust is important
5. Open development is necessary
6. Give credit for digital creation
7. Attitudes must be changed

We summarise the responses from the break-out groups under each of these categories below:

NOTE: In a few cases we have inserted an odd italic word or two, followed by a question mark. This is intended to clarify the point we believe was being made by the respondents.

5.4 There is no single common e-Infrastructure

- 1) You can't have a common single infrastructure
 - There are a wide range of users each wanting different sets of things from an e-Infrastructure
 - Each thing will have different sets of users, but no one component is universal
 - Each user has their own perception of what comprises e-Infrastructure
 - A solution that does 80% of the job well is required
- 2) Don't wait for standards, drive them
 - Design, implement, work with what you have
 - Learn from running pilots and then design a more top down system
- 3) Reinvention is part of the process, systems are under continuous development
 - Incremental steps or leaps of progress (*announced?*) at each AHM?
 - Want a more routine interaction with developers
- 4) You should be able to use packages wherever you like, not just on one machine

5.5 Ease of use is the initial barrier

- 1) Usability is the primary issue for uptake of e-Infrastructure services and the research done using the resources
- 2) Adoption requires ease of use
 - low barriers = large wins
 - Grids should be invisible to end users
- 3) Usability implies usable by users themselves
 - Tools should be adaptable to users' needs

- Having a user base amongst collaborators and education will help demonstrate (*utility?*)
- 4) Target tools that people are using now and adapt these to use e-Infrastructure
 - Opt out rather than opt-in
- 5) Target people developing tools across research areas to how they might target grids
 - commercial packages have established user base
 - Traditional standards process can be too slow for this

5.6 Dealing with complexity is complex

- 1) Where are the boundaries between the levels of complexity that different users will see?
 - Hide grid details from novice users
 - Allow power users to see more
- 2) Security should be appropriate to the resources being used
 - Availability versus identity issues with data (e.g., particle physics vs. medical data)
- 3) The more resources that are available, the harder it is to choose between them
 - Information services are vital, as is good curation

5.7 Trust is important

- 1) Increasing uptake and usage from just by early adopters to uptake and usage by the mainstream majority is about increasing potential users' trust in the infrastructure
 - Trust that knowledge (*generated using the infrastructure?*) can be reused
 - Trust that services of the infrastructure will continue to be provided over research project timescales
 - Trust that data storage will match requirements (*for archival and accessibility?*)
 - Trust that the 'Executables' will create dependencies – will they be around for the length of time that the workflow will be around?
 - Trust in redundant storage sources
 - Secure in terms of knowledge content – Service provider should be disinterested in the content
 - Evidence of this trust is driven by perception of continuation of funding and support

5.8 Open development is necessary

- 1) Development is required in many areas
 - Functionality
 - Documentation
 - Training and education
 - Community building
- 2) Need to facilitate collaboration, publication and communication
 - Both in research enabled by e-Infrastructure and in e-Infrastructure development itself
 - Tools should be open, with documentation
 - Data should be accessible, and documented

5.9 Give credit for digital creation

- 1) Adoption will be hastened if there is a benefit to using e-Infrastructure for more than just selfish compute

- 2) Researchers are entirely focused on published articles because this is what gets credit from grant funding bodies
 - No account taken for other contributions
 - Assign credit for contributions to lab notebooks, datasets, blogs, etc.
 - Encourage tools, practice, training and trust to originate records in digital format

5.10 Attitudes must be changed

- 1) There are chasms to be crossed
 - Infrastructure providers find it hard to attract researchers
 - Researchers find it hard to get to use the infrastructure
 - e-Infrastructure is not just about the grey boxes
- 2) Cultural change does not occur overnight
 - Personal communication (*word of mouth?*) is necessary for adoption
- 3) It's not just about the researcher, but also about the sysadmins providing the infrastructure
 - Make sure they don't just say no when asked by a researcher

6 Summary, Conclusions and Next Steps

The responses accord well with the preliminary findings from the interviews conducted by the ENGAGE project; namely that people will tend to prioritise ease of use, support and continued development over a complete feature set. This requires a sustainable community around the software and trust by the users in the e-Infrastructure providers (and vice-versa).

7 Acknowledgements

We record a special note of thanks to Jean-Claude Bradley and Shantenu Jha for so willingly being put “on the spot” and for giving us thought-provoking position statements to help set the context for the discussion that followed.

We wish to thank the organisers of the All Hands Meeting 2008 for allowing us to run this BoF session and all those participants in the AHM who turned up and contributed so willingly and enthusiastically to the session.