

Enhancing Community Intelligence for e-Science

Alex Voss¹, Rob Procter², Marzieh Asgari-Targhi², Bruce Beckles³

¹School of Computer Science, University of St Andrews

²Manchester e-Research Centre, University of Manchester

³University of Cambridge

Introduction

The recent international review of the UK e-Science programme (Atkins *et al.* 2009) has provided a unique insight into the achievements not only of this specific programme but has also provided insights into the state of the development of e-Research methods and e-Infrastructures more generally. However, such a review provides only a snapshot and tends to emphasise the broad policy implications rather than drawing out specific findings that can be brought to bear on e-Research practice more immediately.

Similarly, the projects in the recent JISC-funded community engagement programme have provided insights into the state of adoption of e-Research methods and e-Infrastructure services. All three projects (e-Uptake, eIUS and Engage) have conducted interviews with researchers and have worked together to ensure a consistent consent process, allowing the sharing of interview transcripts and, potentially, an analysis across the three different projects. At the same time, others such as a group as the universities of Cambridge, Oxford and Reading (cf. Spence *et al.* 2010) are also conducting interviews with researchers for the purpose of informing service provision as their institutions. In addition, the UK's National Grid Service and other service providers conduct surveys of their populations of service users, similarly, with the aim of improving service delivery.

There is by now a significant corpus of evidence available that could further inform and shape the development of e-Research as well as ensure that the impact this programme has on research practice is made evident to funders, addressing important concerns about sustainability and the effective exploitation of investments made. What is problematic, however, is that the material is generally available only to the research groups that have generated it and that effort to exploit its full potential is often unavailable. The differences in methodologies used further compound the situation, putting a limit on the possibility to conduct analysis across the whole aggregate corpus, often limiting what is possible to rather weak [post-hoc comparisons].

The lessons learned from the JISC community engagement programme and similar initiatives over the past years have shown the value of research into the uptake and usage of e-Research methods and e-Infrastructure services. Recent work conducted by some of the authors to re-analyse the material collected in the e-Uptake project has teased out a rich set of specific recommendations for different stakeholders (service providers, institutions and funding organisations). However, analyses of individual studies that capture snapshots will not provide the level of evidence required to underpin the increased community involvement recommended by the International Review of UK e-Science (Atkins *et al.* 2010, p. 32).

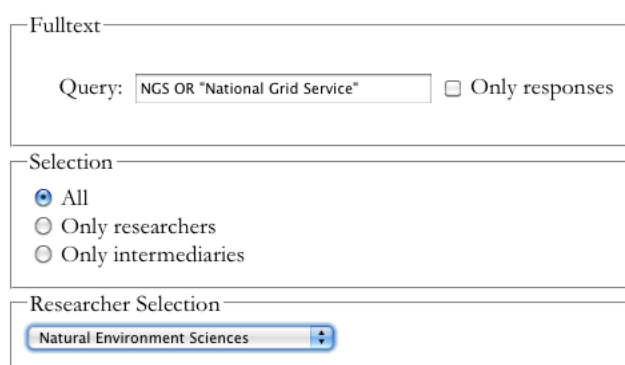
Towards Community Intelligence

There are two aspects to what can be done to move forward from this point. The first is to seek to maximise the value of existing datasets by enabling their further analysis, either individually or in conjunction with each other. The second is to draw up a framework to enable future work to be conducted in such a way that the evidence gathered is complementary and at least broadly comparable as well as to ensure that the subsequent analysis is conducted in a way that ensures

the early participation of a range of stakeholders to bring to bear the wealth of expertise that exists in the community and to ensure that lessons learned are immediately operationalised. We will address each of these two aspects in turn and describe steps we have taken to work towards a better approach to collecting community intelligence and inform future community engagement activities.

Exploiting Existing Evidence

As part of work within the e-Uptake project, we have developed the concept of a 'database of findings' (first described in Voss *et al.*, 2009). Within the project itself we have produced two prototypes focusing on the raw transcript data and our analysis of barriers to uptake and enablers respectively. As part of our activities to follow on from the e-Uptake work, we are currently developing a system that aims to synthesise the two components and to also enable the incorporation of data from future studies. The system currently contains all data from the e-Uptake project and we are developing query interfaces to narrow down the search by subject area and interviewee role (researcher vs. intermediary) and to conduct a full-text search using the Lucene search engine.



The image shows a search interface with three main sections. The top section, labeled 'Fulltext', has a search box containing the query 'NGS OR *National Grid Service*' and a checkbox for 'Only responses'. The middle section, labeled 'Selection', has three radio button options: 'All' (which is selected), 'Only researchers', and 'Only intermediaries'. The bottom section, labeled 'Researcher Selection', has a dropdown menu currently showing 'Natural Environment Sciences'.

Figure 1: Search interface for the database of findings

Our next step will be to re-consent study participants from the Engage project in order to enable us to include the existing transcripts to be added and a first analysis to be conducted making use of the corpora from the two studies. The incorporation of additional interview material from other studies such as the Cambridge, Oxford and Reading ones will require the production of full transcripts as well as the re-consenting of interviewees to incorporate these corpora as well.

Informing Future Studies

In addition to the substantive outcomes, the community engagement projects funded by JISC and the studies conducted at Cambridge, Oxford and Reading have produced important lessons about the practicalities of community engagement and the collection of community intelligence. It is worth capturing this experience in a form that allows it to be reused (potentially by other parties) and to enable future studies to collect data in a way that enables its subsequent curation and analysis across different corpora.

For example, common standards for coding features in the interview transcripts such as barriers and enablers can enable searches across different corpora. A standard consenting process can help to avoid problems such as lacking consent for future use or for sharing of anonymised extracts in the wider community. The production of quality transcripts is costly and error prone, so practical advice on measures to make it effective and perform adequate quality control can help ensure costly mistakes are not repeated and funding is used effectively. One problem observed in the e-Uptake project is that transcribers are often not familiar with the names, acronyms and technical terms used in e-Research, so systematic mistakes are often made that can be avoided if a glossary of terms is provided up front. Finally, as discussed above, a common repository of evidence and formulated findings can help to ensure that best use is made of the material generated.

The workshop at the 2010 UK e-Science All Hands Meeting aims to bring together a wider range of stakeholders with experience in community engagement. We will report on the outcomes of this workshop in a full version of this paper.

Conclusions

We have outlined our work following up from studies of uptake and usage of e-Research methods and e-Infrastructure services. Our aim is to inform future community engagement activities aiming to enable more researchers to benefit from methodological innovations in their disciplines and improved practices utilising advanced information technologies. We have outlined some of the challenges in making effective use of the wealth of material collected by previous studies and in ensuring that we conduct future studies not to provide additional but disconnected snapshots but rather constructs them as parts of an ongoing programme of community intelligence that combines specific funded studies with the everyday gathering of evidence undertaken by practitioners such as service providers and members of research computing services. We believe that if this programme is adequately resourced and carried out with the support of the wider e-Science community it will generate the evidence required for informed policy making and the making of strategic decisions by institutions, funders and technology providers, that it will help to formulate the requirements for the development of new tools and services and, in the end, will lead to better quality research and an increased impact in society.

References

- Atkins, D. *et al.* (2010) Building a UK foundation for the transformative enhancement of research and innovation: report of the international panel for the 2009 review of the UK research councils e-science programme. Swindon, UK: RCUK. See <http://www.epsrc.ac.uk/research/intrevs/escience/Pages/townmeeting.aspx>.
- Spence, Beckles and Martinez (2010). What do we want?: reports from practising researchers in three UK universities. Unpublished manuscript.
- Voss, A., Asgari-Targhi, M., Procter, R., Halfpenny, P., Dunn, S., Fragkouli, E., Anderson, S., Hughes, L., Fergusson, D. & Atkinson, M. 2009 Adoption of e-Infrastructure Services: inhibitors, enablers and opportunities. In *Proc. 5th int. conf. on e-Social Science*, Cologne, June. See <http://www.merc.ac.uk/sites/default/files/events/conference//2009/papers/Voss.pdf>
- Voss, A., Procter, R., Asgari-Targhi, M. and Fergusson, D. (2010). Adoption of e-infrastructure services: configurations of practice. *Philosophical Transactions of the Royal Society A*. (forthcoming)