The University of Manchester Research

e-Infrastructure Development and Community Engagement

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Voss, A., Mascord, M., Arguello, C. M., Asgari-Targhi, M., Procter, R., Fraser, M., M, U., Jirotka, M., Halfpenny, P., Fergusson, D., Atkinson, M., Dunn, S., Blanke, T., Hughes, L., & Anderson, S. (2007). e-Infrastructure Development and Community Engagement. In *Third International Conference on e-Social Science*

Published in:
Third International Conference on e-Social Science

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

OPEN ACCESS

Download date:08. Jun. 2022

e-Infrastructure Development and Community Engagement

Alex Voss[1], Matthew Mascord[2], Mercedes Argüello Casteleiro[1], Marzieh Asgari-Targhi[1], Rob Procter[1], Michael Fraser[2], Marina Jirotka[2], Peter Halfpenny[1], David Fergusson[3], Malcolm Atkinson[3], Stuart Dunn[4], Tobias Blanke[4], Lorna Hughes[4], Sheila Anderson[4]

[1]National Centre for e-Social Science, University of Manchester
[2]Oxford e-Research Centre, Oxford University
[3]National e-Science Centre, University of Edinburgh
[4]Arts and Humanities e-Science Support Centre, King's College London

Email address of corresponding author: alex.voss@ncess.ac.uk

Abstract. This paper presents preliminary findings from two projects that aim to widen the uptake of e-Infrastructures for research. Through the development of a corpus of evidence and through community engagement, we aim to uncover barriers to adoption, enablers that may facilitate uptake and good practice that may be used by those wishing to engage with e-Infrastructures as a 'beaten path' or pattern of adoption.

Introduction

The academic research community is entering a period in which a new kind of digital infrastructure comprised of distributed, networked, interoperable computing and data resources is becoming widely available. Commonly known as cyberinfrastructure in the US and e-Infrastructure in the UK, it is expected to lead to new forms of research (sometimes referred to as e-Research), including enabling and promoting large-scale and interdisciplinary collaborations that, over time, will become accepted, essential components of research practice across all disciplines. In order to capitalise on these opportunities, the pathways for the diffusion and adoption of infrastructural innovations must be first understood and then made as smooth and well supported as possible. In the UK, the Joint Information System Committee (JISC) has funded two projects that aim to study barriers to uptake as well as ways in which e-Infrastructure is adopted in order to develop interventions that will facilitate uptake and make e-Infrastructures available to a wider group of researchers. This paper gives an overview of previous work done in the area and the results achieved so far by these two new projects.

Patterns of adoption

It is sometimes possible to identify patterns of adoption as people apply roughly similar approaches to evaluating the potential usefulness of different technologies. For example, people may tend to follow the example of leaders in their community or may tend to wait until commercial offerings are available. For some users, the 'network externalities' – where the value of a product or service increases with the number of users (Callon, 1994) – e-Infrastructure exhibits may be important. Late adopters may also be less technically skilled or more risk averse, preferring to wait for technologies to stabilise, for entry costs to fall and to capitalise on the experiences of early adopters (Williams *et al*., 2005). Understanding these patterns gives us a handle on what interventions might increase the uptake of e-Infrastructure and lead to more sustainable use. Strategies need to be devised to address barriers, for example, by providing clear development roadmaps and migration routes. Late adopters, for example, may require direct and personalised support in the form of staff development courses (both face-to-face as well as supporting self-paced and remote learning); specific consultancy to develop new applications to utilise services in novel ways; and a single well-curated source of exemplars and information about technical components and services. Different user communities will be at different phases of the adoption cycle at any one time and so support has to be provided for all phases simultaneously. As communities' requirements mature, their support needs may also change. e-Infrastructure users will go through cycles of evaluating requirements and assessing the appropriateness of services while providers will similarly go through cycles of improving services and developing new ones to meet developing needs.

A recent report has highlighted the fact that while any specific process of infrastructural development is contingent and while there is much uncertainty at any point in time, there are also "shared patterns, processes, and emergent lessons that hold widely true across the comparative history and social study of infrastructure" (Edwards *et al*. 2007). Two related insights are that "effective infrastructures are rarely 'built' in an entirely top-down, orderly, and blueprint-like way" (*ibid*., p.2) and that use of technologies, and in particular infrastructural ones, is often deeply embedded in a complex web of socio-material relations.

Innovations in infrastructure succeed because they are able to mobilise a socio-technical constituency (Molina 1995, 1997) – an alignment of technical components, standards, etc. (the 'technical infrastructure') and stakeholder interests, working practices and organisational structures etc. (the 'social infrastructure'). In the case of e-Infrastructure, it is important to understand just where technically-led changes will impact on or be resisted by the ways in which research is socially organised (Edwards *et al.* 2007, p.5). Social organisation and technological development *mutually shape* (Williams and Edge 1996) each other, so we need to embrace the notion that we are making targeted interventions in a field of complex relationships and that we can never fully predict their outcomes. History suggests that the most successful attempts to create infrastructures have been undertaken by those who understood this and have consequently sought to align their technological projects with interventions in the social space (Hughes 1983).

The disconnect between the drivers of a technological vision and the various 'users' of these technologies has been a problem in IT for a long time. In e-Research, this is emphasised by the fact that it involves multiple, largely independent stakeholders with their own separate agendas which are only partially and temporarily aligned and by the fact that the aim of the exercise is not merely the automation of an existing process but a radical socio-technical innovation of research practice. Visions of why this practice should be transformed, just how and using what means are likely to differ. Finally, we need to consider ownership and investment models, or what Edwards *et al*. (2007) call the 'political economy of infrastructure'. At the highest level there are questions about the role of public funds, central steering versus decentralised development and decision making, and (trans-) national cooperation and coordination. At the project level, issues of confidentiality, intellectual property, commercial exploitation of results and etiquettes in collaboration are of importance.

In order to achieve wider uptake of e-Research and sustainable use of e-Infrastructures, it will be crucial to drive developments *within* research communities rather than just offering services *to* them from an outside position. Ultimately, the vision behind e-Research will only be accepted, taken up and further developed by researchers if it becomes their own project and is shaped by their own concerns. This is the aim of two JISC-funded projects to which we now turn.

JISC e-Infrastructure Programme

As part of its mission to provide world-class leadership in the innovative use of information and communications technology to support education and research, the UK Joint Information Systems Committee (JISC) has created an e-Infrastructure programme to expand the uptake and effective use of e-Infrastructure from early adopters and researchers across disciplines. Its objective is to push e-Infrastructure forward in key development areas, to broaden use of e-Infrastructure within the research community including the sciences, medicine, arts and humanities and social sciences as well as to increase capability, expertise and effective use of e-Infrastructure.

The Enabling Uptake of e-Infrastructure Services project (e-Uptake) and the e-Infrastructure Use Cases and Service Usage Models (eIUS) project are funded under the Community Engagement strand of the JISC e-Infrastructure programme. e-Uptake is a collaborative project involving the UK National Centre for e-Social Science (NCeSS), the National e-Science Centre (NeSC), and the Arts and Humanities e-Science Support Centre (AHeSSC). Its aim is to develop strategies aimed at increasing and widening adoption of e-Infrastructure and significantly increasing the user-base of JISC-funded services. The eIUS project is led by Oxford University's Research Technologies Service and e-Research Centre in partnership with NCeSS. Its aim is to gather and document evidence of how e-Infrastructure is currently being used, or how researchers are planning to use it to facilitate the research process (across all major disciplinary areas).

The underlying assumption throughout in both projects has been that technical and social issues are not separate but that there is a mutual shaping dynamic between them. The JISC e-Infrastructure programme will succeed only when social, organisational, and cultural issues are resolved in tandem with the creation of technology-based services (cf., Edwards *et al.* 2007). Having started in April 2007, both projects are work in progress. In the following, a sketch of the study design, research methods employed so far, and preliminary findings will be discussed.

Methodology

A series of case studies are being conducted through surveys of research communities, semi-structured interviews and focus groups with key stakeholders, including: researchers, e-Infrastructure builders, members of e-Infrastructure support initiatives, resource providers and funding agencies. The study design incorporates several key dimensions: discipline (physical sciences, systems biology, medicine, social sciences, arts and humanities), e-Infrastructure components (e.g., middleware, security, service registries, portals) and services, for example, the UK National Grid Service (NGS), Digital Curation Centre (DCC), Access Grid Support Centre (AGSC) and the National Centre for Text Mining (NaCTeM).

The complementary aims of the two projects, the involvement of NCeSS in both and the practical issues raised by conducting research within a relatively small community have encouraged us to undertake a joint approach to the design of the data collection phase. So far, we have conducted a set of pilot interviews involving a dozen researchers from a range of different disciplines and backgrounds. An important aim of this initial period is to refine the data gathering methods and inform the development of our analytical approaches.

We conducted a literature review with the intention to survey different approaches and review previous findings documented in reports, project documentation and the academic literature. These documents provide strong evidence of the advantages of using e-Infrastructure in some application domains. However, the literature reviewed also highlights that the field has not reached the necessary level of maturity and that significant challenges remain in many areas. The lack of maturity in the field is demonstrated by the lack of a standard vocabulary of well-defined terms and the lack of an accepted literature that would aid newcomers to the field[1].

Reflecting this immaturity is the question of what services are associated with e-Infrastructures for research and should be included in the scope of our studies. In scoping the selection of services we would investigate, we decided to include only services that are mainly research specific, provide advanced functionality and are widely applicable, i.e. not specific to a single project or location. Note that scale was not one of the criteria as many important issues in the social sciences and arts and humanities do not involve datasets of a similar size to those dealt with in, say, particle physics or astronomy (the challenges lie mainly in the complexity of the data, not its raw size). For example, we would not consider generic desktop applications such as text editors or spreadsheets to be sufficiently specific to e-Research while we might include statistical packages, especially where they are linked to back-end resources such as data or compute facilities.

e-Infrastructures support activities throughout the whole research lifecycle spanning from the birth of an idea through its recognition by being funded, getting worked on and validated by researchers, results being published and used in further work. There is no one single way of describing the cycle and in practice the picture may be more complex, but we find it useful to work with this slightly idealised picture to structure our findings. The starting point in any research effort is an initial research idea or hypothesis that will lead to the creation of a consortium of researchers writing a project proposal. This involves elaborating the problem and developing research strategies. If the proposal gets the go ahead from funding bodies, there will be a phase of negotiation to work out the detailed division of labour and a project plan. Project execution often starts with a literature review followed by data collection, analysis of the data collected, and discussion of the findings. Collaboration between researchers is a feature of most of these activities although the extent to which activities are undertaken collaboratively may vary. Finally, dissemination of findings through publication and preservation of scientific findings is crucial for the sustainability and follow-on of the scientific work.

Of course, there are further distinctions that will help us scope our study and structure our findings. Research data comes in different types – primary empirical data, secondary data, experimental data as well as data produced by simulations. Different modes of analysis are

---

[1] Most relevant books are edited collections consisting mainly of project reports that are only very loosely connected.

employed in different disciplines and can range from quantitative analysis using statistical methods to qualitative methods to critical analysis applicable in the arts and humanities.

Our close links with the service provider community play an important role in approaching researchers in the evidence-gathering phase. Researchers who are already members of communities such as the NGS User Forum will be relatively easy to identify and approach. This joined-up approach also allows service providers to benefit more directly from our activities and to feel results directly into their service development efforts. While the object and scope of the two studies is the same and they employ the same data gathering methods, the analyses and the outputs produced will differ. While e-Uptake will focus on reasons for non-adoption as well as enablers and interventions to overcome them, eIUS will mainly document existing successful practices. We will discuss the analytic approaches and outputs in the sections below.

The e-Uptake Project

The e-Uptake project involves a series of studies to map the adoption of e-Infrastructure across different research fields, and investigate similarities and differences between them. It will then use the findings to identify the main barriers to and enablers of wider adoption of e-Infrastructure and how they manifest themselves within different user communities. Following from this, we will identify appropriate technical and non-technical responses to these barriers. The project deliverables will have a direct value in helping to inform service providers in their future development of services for these communities. They will also cultivate and influence the provision of training to address the identified barriers and help to shape the development of specific support services and self paced support materials to provide long-term help for these communities. A UK 'One-Stop-Shop' of information about training, outreach and education events, support contacts and available training and support material will help consolidate and disseminate existing offerings as well as those produced by the project itself. In addition, funding agencies will be able to tune their future calls in response to project findings.

Initial Findings

We will now turn to some of the initial findings from our literature review and the interviews we have conducted as part of the scoping phase of the two projects, focusing on barriers to uptake first; more findings will be presented below in the description of the eIUS project.

Researchers in all disciplines have to the deal with the phenomenon of the "data deluge" (Hey and Trefethen 2003, Lyon 2007). Every discipline investigated had its own particular method of gathering, storing, processing and analysing data. Almost all the interviewees acknowledged the use of some form of e-tools in their respective work. The question of access to and management of data emerged as a core issue in data gathering. For example, informants consistently stated that there was no uniform method employed in storing data among the researchers, even within the same research group. Such idiosyncratic practices can lead to problems in scientific collaborations but also have a significant impact on the curation of research results and their reuse (cf. Lyon 2007).

In the stage of data processing and analysis, researchers' choice of e-Infrastructure and tools were complex. It is fair to say that first and foremost they used techniques or strategies that they were familiar with, that were recommended by their peers and colleagues, or that they could learn in a relatively short space of time through training and workshops. There is currently little evidence of a systematic consideration of e-Infrastructures and tools as part of the overall scientific approach. Tool use can also be hindered by other factors such as lack of internationalisation. Researchers whose area of research involved the use of a language other than English often use only those scientific applications that would not require translating the original data into English first. For instance, researchers working on studies of the built environment reported that they conduct their analysis manually, that is, using various types of annotations, physical arrangements of data, etc. or using (older) versions of applications that were available in the appropriate language.

In terms of the use of new e-research tools, researchers were, in general, unhappy about the support or guidelines written by computer scientists. In particular, documentation is often written in a language familiar to and used by computer scientists but that is unfamiliar and incomprehensible to researchers from other areas. A related issue was raised by an evolutionary biologist who was a comfortable e-tool user and programmer in Java. The fact that he wrote his code in Java made it difficult for him to use compute resources that catered primarily for users of C/C++ or FORTRAN. Their experience, furthermore, was that there were issues with establishing a common ground with support staff or those providing training in terms of the use of computer science concepts and terms that researchers may be unfamiliar with. This lack of common ground can lead to situations where researchers felt that they were treated with contempt because of their ‘lack of knowledge’.

Confidentiality of data and security were important considerations and were raised throughout the interviews. In particular, using UK census data proved problematic for an econometrician. The envisaged confidentiality restrictions for the publicly accessible version of the 2001 data meant that variables required for the analysis were not available in the dataset. The project opted to use the 1991 dataset instead. The restrictions of the variables are lifted in the ‘controlled access’ version of the 2001 data but the access constraints on this data mean that it is not available for processing on grids (cf. Peters *et al*. 2007).

The rate of innovation and the bewildering range of options that researchers face when adopting tools can lead to issues with adapting tools and infrastructures because researchers have only limited time for (re-)training. The lack of standardisation of tool usage within disciplines can lead to this being foregrounded, for example, when people move from one research group to another. Usage of different, incompatible tools can also hinder cooperation between researchers from different disciplines or organisations. Achieving sustainable use of e-Infrastructures and tools in the face of constant change and limited time available for appropriating new developments is a challenge. This can force people to make the decision not to adopt a new tool or infrastructure even when there might be obvious benefits.

Another interesting point raised was that e-Infrastructures could narrow down the research questions rather than enlarging its scope. For instance, corpus linguists who employed a particular application in their work tended to be interested in questions that this particular tool could provide answers to, rather than focusing on wider research questions. Respondents commented that this was considered as a barrier to progress in their field.

Researchers are often regular users of electronic communication services and, depending on the context, they employ various kinds of technologies to communicate and exchange data in a number of specialised formats. How researchers collaborated was an important question addressed in the interviews. The issue of trust was raised once or twice but in general researchers thought trusting their peers form other universities was the basic building block in establishing a healthy and productive relationship and in most cases they were using various e-research tools to collaborate. A corpus linguist mentioned how she had published a paper with someone whom she had never met before. They communicated via email, then met at a conference to merge their texts and published a joint paper. However, researchers had mixed views as whether conference tools were useful or not, for instance Access Grid was useful to some but was seen as being time consuming by others.

The eIUS project

The eIUS Use Cases and Service Usage Models project is not simply a requirements-related exercise but rather is intended to broaden participation in the use and future development of e-Infrastructure services. In this respect, the project builds on work done by early adopters in order to appeal to 'mainstream' and 'late' adopters. The project is gathering evidence of how e-Infrastructure is currently being, or genuinely planned to be used. The evidence collected will be documented in the form of *experience reports*, designed to play a pivotal role in the subsequent construction of *use cases* and *service usage models*. Use cases are a semi-formal technique developed in software engineering to express the functional requirements of a system, in particular, how via a sequence of steps, a user interacts with it to achieve a specific goal. In eIUS, the use cases developed will be no more than idealised 'stories' or scenarios that show how researchers currently use e-Infrastructure to move forward with their research agenda. The aim is to tease out patterns of technology use that can be used to provide an index to the rich material assembled in the experience reports (cf. Martin *et al*. 2001). Service Usage Models (SUMs) are a new concept developed by the International e-Framework Initiative, funded by JISC and the Australian Government Department of Education, Science and Training (DEST). Their aim is to show how the scenarios may be realised through the combination of existing and future planned e-Infrastructure services.

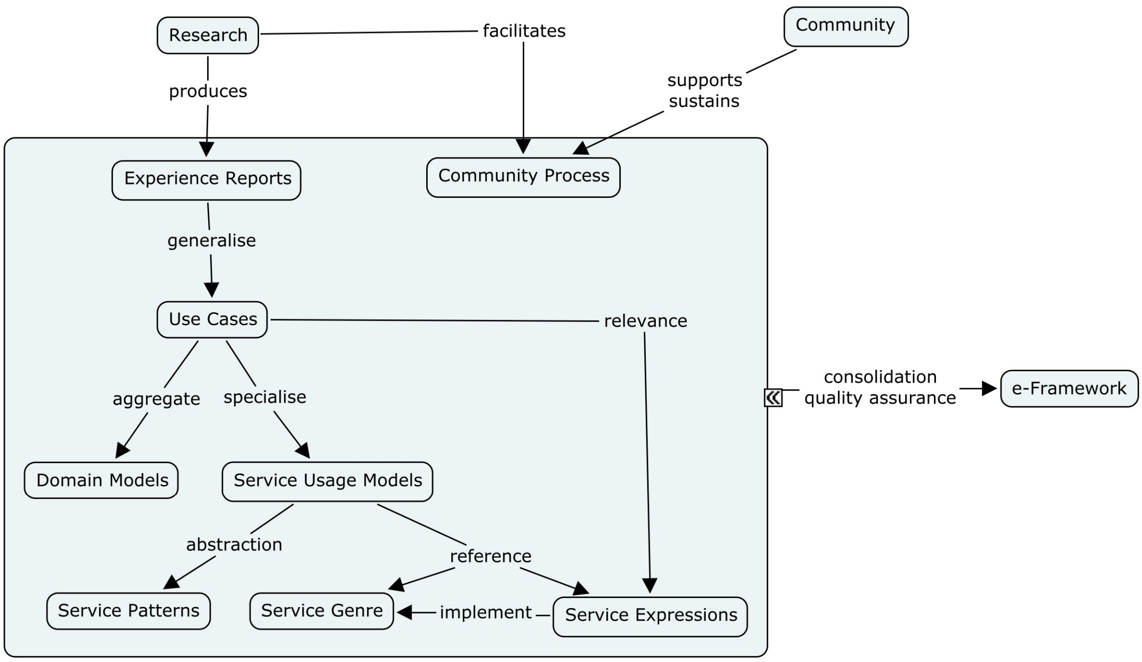

Figure 1: eIUS Evidence Gathering and Modelling.

Each use case, which will be linked back to a series of experience reports, will be developed incrementally through a process of frequent review and validation. A key emphasis in the eIUS project is on *traceability* in the sense that the eIUS project intends to provide the means to allow users to go from the idealised 'stories' all the way back to the experience reports giving specific and concrete examples. Multimedia such as video vignettes of researchers using e-Infrastructure will also be incorporated, where appropriate, to offer to the research community a more meaningful interaction with the use case and avoid a simple document-centric approach. The relationship between the experience reports and the different models based on them can be seen in Figure 1.

The Community Process

For the reasons stated above, the outputs produced by the eIUS project in the form of Experience Reports, Use Cases, and SUMs are anticipated to be useful to service providers and funding agencies as standalone entities in their own right. However, a potentially much more valuable outcome is the creation of a *community* that can ensure the ongoing sustainability of this kind of discipline-wide information gathering and engagement activity far beyond the project's formal end date. In a consultation with the UK's OSSwatch[2] advisory service, the idea of seeding such a community was explored in relation to the types of communities that are frequently created to support Open Source Software development projects.

In successful Open Source software projects, communities of users, contributors and committers form around software to support and develop it into the future. The vast majority of community members are simply users and are largely inactive. Some members may encounter problems when using the software and submit a bug report, at which point that they move into the next level of involvement as *contributors*. Even more active members, willing to invest their institution's or their own time may develop and may submit a *patch* that fixes a bug, or develop a new feature. The most active members may eventually be accepted as committers, at which point they take on joint responsibility for the project's outputs.

Relating this to the eIUS project, one might imagine that while clearly the researchers themselves are ultimately the intended audience for the use cases and experience reports, they may not be the most active members of the community developing them. The immediate addressees of the community development effort are therefore more likely to be e-Infrastructure providers, who will also act as moderators between the project, the community process and the research communities. The formation of any community is hard work, difficult and success is contingent on a number of factors beyond the project's control. Furthermore, the overheads of supporting this community once established should not be under-estimated. As the formation of a community cannot be a straightforward project deliverable, it is all the more important to ensure that the project's efforts are linked up with existing programmes and communities so as to complement and add to them.

Clearly, the involvement of researchers is absolutely critical to the success of the project. However, we feel that the engagement of researchers can only be achieved indirectly through the creation of the community of service providers. Researchers are not going to see the benefit in engaging with yet another community once removed from the user communities associated with the e-Infrastructure services they are already engaged in. In contrast, service providers *will* see the benefit in engaging with the community in terms of pooling resources to do requirements analysis, user engagement, education and outreach, which it is hoped will lead to increased take-up and realisation of the full potential of e-Infrastructures.

---

[2] www.oss-watch.ac.uk

Having said this, an important aspect of building a community is to provide space for participation that will allow people to contribute and submit items even if they are unlikely to do so. The community portal will provide a location where the experience reports, use cases, and service usage models, can be deposited, peer-reviewed, consumed, and further enhanced. The eIUS project differs from previous projects in that it sets out to collect concrete information about use of e-Infrastructures for research and to make it available in an easily accessible form. Our initial findings suggest that researchers are motivated to participate in the study and might potentially engage with the community process for a number of reasons. These range from simple goodwill, phrased as 'giving something back', to the expectation of some tangible benefit, such as highlighting the needs of a particular class of researcher that may have thus far been overlooked.

Initial Findings

Although the study is still at a preliminary stage, with only a dozen researchers interviewed to date, the following gives a feel for the types of experience reports and use cases we expect to emerge through the continuation of this process. From the dozen interviews carried out to date, engagement with e-Infrastructure seems to vary dramatically between research disciplines predominantly characterised by *qualitative* analysis methods as opposed to those characterised by more *quantitative* methods. For example, researchers in the built environment, and educational studies frequently produce and process data from interviews using transcription packages and qualitative analysis tools such as NVivo or Atlas.ti. Many of these tools are predominantly standalone desktop application with little apparent need for connectivity with any external database or service. There are exceptions though, as we found a couple of examples where cognitive mapping tools were used, and these researchers were starting to use web-based tools that allow these cognitive maps to be shared amongst collaborators. Nevertheless, e-Infrastructure's applicability becomes more apparent when considering the more quantitative aspects of social, physical or natural sciences.

Linguistics is one example of an area where e-Infrastructure is becoming increasingly important. Corpus linguistics, in particular, is an area fundamentally allied to quantitative analysis techniques and has moved on considerably from the days when 'hordes of monks' would spend decades analysing texts by hand. Nowadays, large-scale web-based databases of written and spoken language, such as the British National Corpus (BNC)[3] are nearly always used as the basis of this analysis. The interfaces to these corpora usually take the form of web or desktop based *concordance tools* that allow many different kinds of textual analysis beyond simple searches for strings of characters. For example, annotation and sorting on preceding or following words or words some defined distance away. It was reported that very few linguists have the skills required to write their own analysis tools and off-the-shelf tools and services play a critical role in the work corpus linguists carry out.

Condensed matter physics, which concerns the physics of matter at extremely low temperatures, is a much smaller scale activity than particle physics, both in terms of cost and the amount of physical space required. As a consequence, the field is characterised by several small experimental groups in contrast to large, centralised collaborations. There is also a greater level of interplay between the theorists and the experimentalists. This interplay works both ways, with results from the theorists influencing the experimentalists, and experimental data being incorporated into the theorists' work. Theorists in the area can take either a

---

[3] www.natcorp.ox.ac.uk

mathematical or computational based route. The lack of off-the-shelf support for this 'green field' area of science means that those taking the computational route often have to write both the simulation code and the underpinning equations from scratch, with some help from online code libraries such as numerical recipes[4]. Once the simulation code has finally been written, there is then the issue of actually finding the computational resources necessary to actually run the simulations; as the groups are small, it cannot be assumed these resources will be available. Campus grids can play a pivotal role here. One of the researchers interviewed was fortunate in that he had been made aware of a new campus grid project at a major UK research university, at exactly the right point in their research, and has subsequently become heavily dependent on it. On the other hand, a colleague who never had access to such facilities took a more mathematical oriented research path for precisely this reason.

In the preceding two examples, the researchers concerned are participating in research primarily for their own benefit, the end-result being papers in refereed journals where they will be high up on the author list. Collaborators, in these cases, will be more or less on an equal footing. In another style of research, the researcher may take a back seat, and effectively provide a *service* to another, rather than leading the research itself. This is often the case with researchers who are experts in a particular analysis technique applicable to a host of disciplinary areas. In the field of medical image analysis, an offshoot from engineering science, for example, researchers collaborate with clinical researchers who have access to medical images. In this role the engineering science researcher is an 'engineer for hire', with the ultimate goal taking the form of a paper in a clinical journal where they will be a co-author, but certainly not the first author. The collaboration can also work the other way round, with the output of the collaboration taking the form of a paper in an engineering journal where he will be first author and the clinician providing a source of 'test' images. This type of research relies heavily on using or adapting existing algorithms and software code, for example, those provided through online code libraries such as the Insight Toolkit (ITK). However, not all journals mandate the submission of code which is vital if the researcher is to validate the code's suitability for use. Another issue is access to test images which cannot always be made available due to the difficulty of completely anonymising medical images.

Figure 2 illustrates what a use case might look like and how it might be linked both to the experience reports (i.e., the evidence base) and to further information such as examples of possible technical choices.

Conclusions and future work

The e-Uptake and eIUS projects are closely related and work in each will inform the other. For example, the evidence base, use cases and service usage models provided by the eIUS project will serve as enablers to uptake and may feed into the production of training materials and events. Our pilot interviews have demonstrated the importance of close cooperation between the projects and we will carry the momentum forward into the next stage of evidence gathering. At the same time, both projects have been engaging with relevant service providers and e-Infrastructure initiatives; this work will be continued and driven forward through a series of workshops over the course of the coming year. Engagement with the user communities will be fostered through training, education and outreach events as well as the development of the community portal.

---

[4] www.nr.com

1. Alex, a researcher working in medical image analysis is phoned up one evening in June by his friend Alison, who produced a sequence of MRI brain images of an unprecedented high resolution from hospital volunteers the day before.

2. Alison, who is interested in researching blood flow through the brain, is hoping Alex can help through the use of image analysis techniques to identify blood vessels, and by analysing sequences of the images estimate the amount of blood flowing through them over time.

> … these images have been acquired at a very, very high resolution, they are sequences of images, so what we want to do is […] we want to calculate how much the volume of the heart changes.

3. Alex decides he is interested in working with Alison as no one has done image analysis on these types of images before and there is potential for research papers both in the clinical journal *Cerebral Blood Flow & Metabolism* as well as in the *Insight Journal*, a medical image analysis journal. They agree over the phone to collaborate and write joint papers on any successful results.

4. Alison sends Alex the images in a DVD through the post, and they arrive in a couple of days. Alex sends Alison an email to arrange a phone conference to discuss the images. In the meeting Alison explains what exactly is shown in the images, and what she wants to get from them.

> Alison could have equally used Secure FTP or Storage Resource Broker (SRB) to send the images. However, her lab had recently introduced a new firewall which had not been opened up yet to allow this and since she was going on holiday for a few days anyway felt it quicker and easier just to burn a DVD and send it by post.

5. After the meeting, Alex does some preliminary analysis of the images on his lab's local cluster using blood flow analysis code that has worked for lower resolution images. As expected, he finds that the contrast in the image has decreased to the point where his existing code is unable to discern the relevant features. Though a major challenge, developing successful analysis techniques for these images has major implications for his career.

6. Following several unsuccessful attempts to develop MatLab code on his workstation to analyse the images, he phones up trusted friend to discuss the problem. His friend tells him about some successful image analysis of low-contrast images of nerve cells, which has been published in the *Insight Journal*.

> When you are dealing with very large data sets, big images, sequences, matlab many times has limitations in terms of memory. Then I go to C++, then If I go to C++ then I will use ITK.

7. Alex downloads this code from the *Insight Journal* which is fortunately supplemented by a series of test images. Over the next three months, Alex works to integrate this code into his existing software, running the analysis on small sections of the image where high performance computing is not required.

8. In the middle of September, Alex finally produces a code that appears to work successfully on small samples and decides it is time to try the analysis on the full sized images. The memory on his desktop will be insufficient to do this kind of analysis so he is in need of his lab's cluster again. To use the cluster, Alex has to do further work on the code to have different compute nodes in the cluster analyse separate chunks of the images in parallel.

9. An analysis conducted on the cluster is successful and Alex contacts Alison to let her know that he is ready to share some data with her. Alison validates the analysed images against her previous manual analysis and is relieved the results are consistent.

10. Over the next two weeks, Alex carries out the analysis on further images, yielding invaluable clinical data, which is used by Alison in a paper accepted by *Cerebral Blood Flow & Metabolism*. Alex writes up his new analysis technique and publishes it in the *Insight Journal*, submitting the associated code and images, ensuring others can reproduce his results in future.

Figure 2: Example Use Case Scenario.

Acknowledgements

This work has been funded by the UK's Joint Information Systems Committee under the community engagement strand of its e-Infrastructure programme. We would like to thank our informants, in particular Dr Vicente Grau, who has helped us create and validate the example use case.

References

Callon, M. (1994). Is Science a Public Good? Science, *Technology, & Human Values*, Volume 19, Issue 4, Autumn 1994, 395-424.

Edwards, P.N., Jackson, S.J., Bowker, G.C. and Knobel, C.P. (2007). *Understanding Infrastructures: Dynamics, Tensions, and Design*. Report of a Workshop on "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. National Science Foundation.

Hey, T. and Trefethen, A. (2003). The Data Deluge: An e-Science Perspective. In Berman, F., Fox, G. and Hey, T. (eds.). Grid Computing: Making the Global Infrastructure a Reality. John Wiley & Sons.

Hughes, T.P. (1983). *Networks of Power: Electrification in Western Society, 1880-1930*. Baltimore: John Hopkins University Press.

Lyon, L. (2007). Dealing with Data: Roles, Rights, Responsibilities and Relationships - Consultancy Report. Available at: http://tinyurl.com/2nknuc

Martin, D., Rodden, T., Rouncefield, M., Sommerville, I. and Viller, S. (2001). Finding Patterns in the Fieldwork. In Prinz *et al*. (eds) Proceedings of the 7th European Conference on Computer Supported Cooperative Work, Bonn, Germany. pp. 39-58.

Molina, A. (1995). Sociotechnical constituencies as processes of alignment: The rise of a large-scale European information technology initiative. *Technology in Society* 17, 385-412.

Molina, A. (1997). Insights into the nature of technology diffusion and implementation: the perspective of sociotechnical alignment. *Technovation* 17(11/12), 601-626.

Peters, S, Clark, K., Ekin, P., LeBlanc, A. and Pickles, S. (2007). Grid Enabled Empirical Econometrics: A Microdata Application. *Computational Econometrics*.

Williams, R., Stewart, J. and Slack, R. (2005). *Social Learning in Technological Innovation: Experimenting with Information and Communication Technologies*. Cheltenham: Edward Elgar

Williams, R. and Edge, D. (1996). *Research Policy* 25. pp. 856-899.