eius
e-Infrastructure Use Cases
and Service Usage Models

www.eius.ac.uk

Scoping e-Infrastructure

Interim Report

Matthew Mascord
Mercedes Argüello Casteleiro
Mike Fraser
Rob Procter
Peter Halfpenny
Alex Voss
Marina Jirotka

October 2007

Table of Contents

Executive Summary

eIUS is an applied research project funded under the JISC e-Infrastructure programme with a remit to create a detailed picture of e-Infrastructure usage across UK academic research and, through actively publicising successful and inspiring use, facilitate an overall increase in take-up. This report positions and scopes the eIUS project in relation to existing initiatives and outlines a fieldwork methodology now intended to be rolled out across the UK.

1. e-Infrastructure, in the context of UK academic research, takes a broad meaning embracing a large number of information and communication technology (ICT) genres and support services. eIUS concerns itself with UK researcher interactions with technologies that support this research and, for the most part, considers only those ICTs that are **supported**, **networked**, **specific to research**, and **applicable to more than one individual or research group**.
2. The term 'use case', in the context of eIUS, is taken to refer to a semi-structured story or scenario showing how researchers today can use e-Infrastructure to achieve specific research goals. It differs from its use in a software engineering context because it is not concerned with the intended user interactions with a future to-be-constructed system, but rather with the totality of e-Infrastructure services as they exist today.
3. eIUS differs from previous work in that it seeks to obtain detailed information on researcher behaviour at the 'coal-face' rather than indirectly through the accounts of e-Infrastructure service providers. eIUS is also relatively unique in that it intends to provide explicit traceability from the use cases all the way back to the evidence collected through interviews and observational studies.
4. Having considered and evaluated a number of different sampling approaches, the project team believes that the best approach for uncovering usage examples and securing the involvement of active UK researchers is through existing e-Infrastructure services. The project team does, however, also recognise the need to dedicate ongoing effort to systematic desk research and to securing the participation of well-connected individuals, to uncover the use of new, emerging and less-well-known services, and services whose development has fallen outside conventional funding programmes.
5. The project team intends to explore the use of more proactive approaches for engaging the so-called 'quiet' users who do not normally respond to conventional calls for participation and to make use of snowball sampling - asking informants who else they know using e-Infrastructure - as another strategy for obtaining a broad and relevant evidence base.
6. In order to facilitate an increase in e-Infrastructure take-up by UK researchers, the project team recognises the need to focus on finding successful and inspiring usage examples that have considerable potential for transferability both within and outside the research domain they originated.
7. Carrying out debriefings at the end of the researcher interviews has proved highly effective in terms of refining the methodology during the pilot study. Asking researchers why they decided to participate has revealed a number of motivating factors that will help the project to effectively market its activities to new researchers in the next phases. These factors include publicity, reflecting on and sharing research methodologies, goodwill (or 'giving something back'), networking, highlighting service problems or issues, representing a specific community need, and empathy with the point of view of the service provider.
8. The project team has found that the level of detail obtained through relatively short interviews with researchers is often insufficient to construct scenarios based on the interview data alone; there is normally a need to carry out additional desk research and to *create* sequences of events providing an overall structure for the use cases. Elements of invention can be minimised through involving the informants in the development of use cases and by asking them to validate the use cases in terms of their believability.
9. The project has developed a working format for the use cases, illustrated by four prototype use cases presented in Appendix A. The team does, however, reserve judgement on a number of editorial considerations that will remain under review in the subsequent phases of the project, and be a topic of discussion at the project's forthcoming community engagement workshops.

1. Introduction

eIUS is one of two complementary community engagement projects funded under the JISC e-Infrastructure programme with a mandate to develop a deeper understanding, and favourably influence uptake of e-Infrastructure throughout UK research.

While eIUS has a remit to encourage uptake through highlighting successful and inspiring use, its sister project, e-Uptake, aims to investigate and address the underlying social, organisational, and technical factors that influence uptake of e-Infrastructure, whether successful or not. The combined outcomes of the two projects are expected to help realise the full benefit of existing e-Infrastructure services in the UK as well as to identify priorities for future e-Infrastructure service development.

This preliminary report presents work carried out to date in the Survey and Scoping e-Infrastructure work package. It outlines the results of a small-scale evidence-gathering pilot that evolved a fieldwork methodology now proposed to be used on a much larger scale within the remainder of the project. The report also positions and scopes the eIUS project in relation to existing initiatives, both past and present.

The intended audience for this report includes the stakeholders of the project who are likely to be interested in the project's initial findings and the fieldwork methodology proposed. These include JISC, who are funding the project, UK e-Infrastructure services that can potentially benefit from the work of the project, and finally, the International e-Framework Initiative who are actively seeking authentic UK research community usage examples.

2. Scope

At the outset of the project there was considerable debate amongst the stakeholders on the Information and Communication Technologies (ICTs) that should come under the banner of 'e-Infrastructure' and thus within the scope of the project. Existing definitions for this relatively new term (roughly equivalent to the US term cyberinfrastructure) were at the time extremely broad and this made it difficult to prioritise the efforts of the project and avoid 'scope-creep'. In order to ensure the most effective use of the project's limited resources, the team developed a set of criteria (illustrated in Table 1) designed to help decide whether a particular ICT should be considered within scope.

	ICTs Within Scope	**ICTs Out of Scope**
1.	Supported	Unsupported
2.	Networked	Standalone
3.	Research-specific	Generally-applicable
4.	More benefit when shared	Less benefit when shared.

Table 1 – Criteria for determining whether an ICT is within scope

The first criterion is relatively weak, in the sense that virtually any usable ICT must require some form of support even if it is merely at the level of providing software updates and technical support. However, the second criterion rules out standalone software where there is no apparent need for connectivity with any form of back-end infrastructure, and the support provided goes no further than this basic level. To be considered within scope, the ICT must be a hosted service accessible over a network, tending to be concomitant with some form of dedicated support that ensures continuity of service. This support does not necessarily have to be provided in any formal sense, but could be provided on an *ad hoc* basis at the research group level, or more formally at a departmental, institutional, national, or even international level (the project is not restricting itself only to the use of UK-hosted e-Infrastructure). Also of less concern is whether open source development communities, commercial organisations or public bodies, provide the support.

Even assuming the first two criteria are satisfied, there are still many ICTs that are of less interest to the eIUS project either because they are very widely applicable and not particularly specific to research, or

at the other extreme, only applicable to a single individual or research group. The third criterion excludes ICTs in the former category, for example, online calendars where usage patterns for researchers are likely to be identical to those found outside academia. Only if the ICT is used in a particularly research-specific way would it considered within scope. Excluding the highly tailored ICTs is the intent of the fourth and final criterion; ICTs with very limited re-usability beyond the context within which they were originally created are out of scope. On the other hand, proprietary ICTs with demonstrable reusability beyond their local context, albeit with a need for further work, are within scope.

Based on these criteria, Table 2 gives a series of concrete examples of ICTs both in and out of scope.

Examples of ICTs Within Scope	Examples of ICTs Out of Scope
Desktop tools specific to research that link or have considerable potential for linkage with back-end resources, for example, data sources, or compute resources. For example, MatLab[1].	Static web sites.
Compute resources of any type: local clusters, high performance computing (HPC), campus grids. For example, OxGrid[2], the National Grid Service[3] (NGS), HPCx[4].	Search engines. For example, Google.
Online journals with associated data repositories. For example, the Insight Journal[5].	Conventional online journals.
Research data archives. For example, the Virtual Vellum Project[6].	Compilers. For example, GCC[9].
Online or networked visualisation tools. For example, VMD[7].	Library catalogues. For example, OLIS[10].
Online or networked data or text mining tools.	Generally applicable ICTs, whether networked or not: web browsers, word processors, text editors, spreadsheets, presentation packages, email clients, calendars, project management tools. For example, Google Documents[11], Microsoft Word, Google Calendar[12], Basecamp[13], Microsoft Outlook.
PDAs when linked with back-end infrastructure, and used in the field, rather than just for generic email or calendaring. For example, the VERA[8] project.	Operating systems. For example, Mac OS X[14].
	Professional or social networking sites. For example, Facebook[15], LinkedIn[16].

1 www.mathworks.com
2 www.oerc.ox.ac.uk/facilities/oxgrid.xml
3 www.ngs.ac.uk
4 www.hpcx.ac.uk
5 www.insight-journal.org
6 www.shef.ac.uk/hri/projects/projectpages/virtualvellum.html
7 www.ks.uiuc.edu/Research/vmd
8 vera.rdg.ac.uk
9 gcc.gnu.org
10 www.lib.ox.ac.uk/olis
11 docs.google.com
12 calendar.google.com
13 www.basecamphq.com
14 www.apple.com/macosx
15 www.facebook.com

Table 2 – Examples of ICTs In and Out of Scope

The other obvious consideration for any project examining the use of ICTs is *who* is actually using them. The eIUS project is concerning itself only with usage scenarios involving active UK-based researchers in any discipline, whilst recognising that national boundaries are frequently traversed by the researchers themselves, either through their interactions with collaborators outside the UK, or through their use of e-Infrastructure services hosted overseas. *Active* in this context is taken to refer to any individual who spends a significant proportion of their work (or own) time carrying out research at the 'coalface'.

3. Previous Work

A literature review was conducted in order to survey different approaches and review previous findings documented in reports, project documentation and the academic literature. The first half of this section provides a summary of the intellectual underpinnings to this work by highlighting relevant work in a number of disciplines. The second half of the section presents examples of recently completed projects that are similar in aims, scope and methodology to eIUS, and shows how eIUS complements and builds on them.

3.1 Patterns in the Use of e-Infrastructure

In describing e-Infrastructure usage, we are interested in uncovering patterns of usage that can be represented in a form that is easy to digest, abstract enough to make the key features stand out yet concrete enough to engage the reader and allow them to relate the content to their own situation. Based on and grounded in the richness of the experience reports the project will gather, we are looking for a form of representation that provides the right amount of detail to give a convincing account of usage without getting lost in contingent details. Researchers in research areas such as software engineering (SE), computer-supported cooperative work (CSCW), human computer interaction (HCI) and science and technology studies (STS) have studied the human, social and political factors that are important elements in the development and implementation of information technologies. In the following, we will briefly review relevant work from these communities, focusing on the forms of representation developed, before discussing their relevance to the eIUS project and its own approach.

Within *traditional* human computer interaction, the focus has been largely on the human-computer dyad and interface issues rather than the wider socio-technical environment in which work takes place. While the importance of the latter is often acknowledged, the representations developed in HCI usually describe it only in very vague terms as a set of influencing factors while being much more specific at the level of the individual worker and their interaction with the machine. The development of pattern languages in HCI also tends to be prescriptive and design-oriented rather than aimed at uncovering and documenting use practice (*cf.* van Welie and van der Veer 2003). However, the boundaries of HCI are hard to define as there are many overlaps with related communities such as CSCW. Indeed, many authors have published their work in both communities.

CSCW studies make the social organisation of activities their central concern, often producing rich descriptions of collaboration in everyday (work) settings. The use of ethnographic studies to inform design has become an established feature in user-centred systems development. However, while ethnographies have played an important role in highlighting the routine features of everyday working lives and the situated accomplishment of activities (Hughes et al., 1992. 1997), the development of requirements based on this understanding has remained problematic (Plowman *et al.* 1995). For example, there is no principled way of moving from a description of how work is organised to how it *ought* to be organised. Also, it is difficult to communicate the findings of the ethnographic study and there is a tension between the richness of ethnographic observation and the need to distil this into a set of concrete recommendations for design. What the approaches described below have in common is an

16 www.linkedin.com

interest in integrating such qualitative empirical data and social analysis into the process of systems development, implementation and use.

One way in which ethnographic studies can inform design is through the use of typifications, either of technologies or of observations. Trigg, Blomberg and Suchman (1999) provide an example of the former through their study of issues arising from a particular class of system, namely document management systems. They provide a set of questions that one might ask in relation to a setting in which such a system might be introduced. These arise from the specific study undertaken but are formulated as questions one might ask of any 'typical' setting where documents are managed, so they are made more generally useful. A related approach is to distil a set of findings into patterns (Martin *et al.* 2001, Martin and Sommerville 2004) in order to make them available as a "background for understanding or characterizing work in different settings" (*ibid.* p.62). By making findings available in a standardised format and pointing out how they are *repeated findings*, patterns aim to make the body of fieldwork data more accessible by linking back to a number of actual studies. The value lies as much in the systematic representation as in the fact that it makes a more substantial resource accessible. Patterns are orienting and organising devices that provide topics for investigation and ways of relating findings to similar ones from other settings, opening up the possibility of comparison and contrasting.

The specific patterns that Martin *et al.* (2001) have collected are repeated observations of the social organisation of work such as 'artefacts as an audit trail' – describing how an artefact such as a paper form can represent the history of their production, through annotations, stamps, attached documents, etc. (*ibid.*, p.53). Each pattern contains a short description, a note on its relevance, possible implications as well as short summaries of at least two studies that informed their formulation, including links to the original studies. It is important to note that in contrast to other uses of the notion of a pattern (e.g., Gamma *et al.* 1994), Martin *et al.*'s patterns are descriptive rather than design-oriented and in this respect are consonant with the use of use cases in the eIUS project as they serve to document and describe common interactions in complex socio-technical settings. On the one hand, they can be used to communicate the results of a specific analysis to designers (e.g., service providers) to be used as a resource for design or service development. On the other, they can provide a way in which important features of observed usage can be highlighted and presented in a manner that is more accessible to a wider community (of researchers).

Software engineering methods make use of a wide range of graphical and textual representations that can be more or less formally defined. Usually, these representations are prescriptive and design-oriented, i.e., they are intended to define aspects of the future functioning of a system to be implemented rather than to describe existing arrangements. The most prominent set of representations is the Unified Modelling Language (UML), a set of diagrammatic representations of different structural and behavioural aspects of the system. As of themselves, the UML diagram types provide little support for representing the rich relationships between technical components and social arrangements that eIUS intends to capture[17].

The Coherence approach (Viller and Sommerville 2000) is an attempt to bring together ethnographic observation, viewpoint oriented analysis and design-oriented representation using the well-established set of UML diagram types. The aim is to improve the way that findings from ethnographic studies can be brought into the design process and be made relevant by connecting them directly to standard software engineering practice and representations readily familiar to systems developers. It focuses on three 'social viewpoints': distributed coordination, plans and procedures and awareness of work. These are complemented by five 'social concerns' that cut across the viewpoints: paperwork and computer work, skill and the use of local knowledge, spatial and temporal organisation and organisational memory (*ibid.* p. 174). Viller and Sommerville show how the viewpoints can be used to elicit specific information about the social organisation of work that directly feeds into UML models. For example, "the distributed coordination and awareness of work viewpoints both identify a number of objects in the domain that play a role in coordinating work between the actors" (*ibid.* p. 181). It is important to

[17] Traditional software engineering methods also do not provide ways to uncover the phenomena that eIUS is interested in but this is not the interesting point here because we wish to focus on the representations.

note at this point that Coherence extends UML models slightly where necessary, for example to represent interactions in use case and sequence diagrams by allowing messages between different actors to be modelled (*ibid.* p. 193). Coherence thus provides a number of 'sensitising devices' and concrete questions to ask during the fieldwork as well a way to link findings into UML representations, some of which can be used to highlight particular features of the work that was studied.

Designing working systems involves moving from current work practice to a new configuration of work involving new technological components, new organisational arrangements or divisions of labour. Traditional supply-driven approaches to IT systems development assume a pipeline model of innovation: systems get developed, rolled out and change working practices. This leads to a consequent assumption "that the primary solution to meeting user needs is to build ever more extensive knowledge about the specific context and purposes of various users into technology design" (Williams, Stewart and Slack 2005, p.67). Stewart and Williams (2005) have described this preoccupation with the initial design of technological artefacts as a 'design fallacy' that neglects the innovation that happens as technologies are deployed and used and as people learn more about their requirements as a result of this (*cf.* Voß, Procter and Williams 2000). A *social learning* perspective (Williams, Stewart and Slack 2005, Sørensen 1996, Stewart and Williams 2005) allows us to describe the innovation process leading to successful e-Infrastructure usage in terms of the complex sets of interactions taking place between a number of different actors. In particular, it allows us to capture the ways in which technological artefacts are appropriated and domesticated, i.e., how they are factored into a stable set of arrangements that allows them to become features of everyday work. This is consonant with the aims of the eIUS project: to document instances of socio-technical organisation that can inspire further uptake and 'innovation in use'. eIUS aims to inspire a form of social learning that Fleck (1988) has described as 'innofusion' (from innovation + diffusion): innovation continues in use as technologies are appropriated and further needs are discovered. These get factored into working socio-technical configurations and into technological design – both of which can be potentially traded between different actors – between users (researchers) or between users and developers (service providers).

3.2 Related Work[18]

The JISC-funded human factors audit of selected e-Science projects (Kalawsky *et al.*, 2006) documents eight UK e-Science projects: MyGrid, eDiamond, GridPP, AstroGrid, HyOntUse, CombeChem, eBank, and RealityGrid. The methodology of the study is based on a previous Human Factors Audit carried out by the same team on the RealityGrid project. The two main sources of the data were project literature and interviews with major stakeholders – both users and developers. Interviews were organized around a questionnaire containing closed and open-ended questions covering: scientific application details, technical environment, task characteristics, user environment, and user types. One precondition for the interviews was that the interviewer would already be familiar with the project, having read the project documentation so as to be able to tailor their questions to the specific project. The project did not set out to develop new use cases. However, several of the use cases generated by the audited projects are themselves included through the project summaries. The report stresses the importance of use cases as a method to specify high-level functional requirements, noting that they were not uniformly used in the various projects but had a positive impact where they were.

The EPSRC and JISC funded SUPER study (Newhouse *et al.* 2007) identifies longer term requirements (3-5 years), as well as outlining a set of short-term (6-18 months) priorities for existing national

18 The project has also considered: Liz Lyon, *Dealing with Data: Roles, Rights, Responsibilities and Relationships - Consultancy Report* (JISC, 2007), http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories/project_dealing_with_data.aspx. Edwards, P.N. , Jackson, S.J., Bowker, G.C. and Knobel, C.P. "Understanding Infrastructure: Dynamics, Tensions, and Design. Report of a Workshop on 'History & Theory of Infrastrucuture: Lessons for New Scientific Cyberinfrastructures'" (January 2007), http://www.si.umich.edu/cyber-infrastructure/UnderstandingInfrastructure_FinalReport25jan07.pdf. Zimmerman, A. and Finholt, T.A. *TeraGrid User Workshop Final Report*. Collaboratory for Research on Electronic Work, School of Information, University of Michigan, (July 2006), http://www.crew.umich.edu/research/teragrid_user_workshop.pdf. NSF Cyberinfrastructure Council, *NSF's Cyberinfrastructure Vision for 21st Century Discovery* (March 2007), http://www.nsf.gov/od/oci/ci_v5.pdf.

e-infrastructure providers, for example, OMII-UK, NGS, and DCC. It identifies five recurring common issues and gives recommendations in three broad areas: software, policy, and support. A first draft of the report, based on interviews with approximately 30 UK e-Science projects, was discussed at a workshop with delegates from the e-Science community in early 2007. The outcomes from this workshop, as well as the results of an online survey are presented as an appendix to the final version of the report, published in April 2007, as a UK e-Science Technical Report (*ibid.*). The methodology used was qualitative in nature, and based upon a series of unstructured interviews in late 2006 with representatives from the e-Science projects. Those interviewed were classified as end-users (current or potential end-users of e-infrastructure), technologists (those building or adapting existing technology for a specific set of end-users), and generic tool developers (those building technology for the 'mass market' for use in multiple disciplines). The findings presented are pitched at a relatively high level and largely avoid citing concrete examples of the actual use of e-infrastructure by discipline communities. In this respect, the study differs from eIUS, which is aiming to provide explicit traceability from the use cases, all the way back to the evidence found within the experience reports.

The EU-funded AVROSS[19] study (Barjak *et al.* 2007) conducted eight case studies of projects relevant in the fields of social sciences and humanities. The projects were identified on the basis of a set of criteria developed from and applied to data from a previous online survey and additional desk research, with the aim to identify cases that hold particular promise. Data about the cases was collected through semi-structured interviews and reviews of project documentation and publications. The interview framework was organised around a number of factors influencing e-Infrastructure development: technological frames and user requirements, scientific shaping of technology, economic factors and political influences. In addition, the interviews aimed to elicit information on challenges and difficulties faced, modifications that were made to initial approaches and experiences with transfer into other environments (*ibid.* p. 75).

The set of case studies contains two that are directly relevant to the eIUS study as they involve JISC funded services. The first describes the use of Access Grid facilitated by the Access Grid Support Centre (AGSC). However, it is largely written from the perspective of the ASGC and does not contain descriptions of use practice. The second describes the MoSeS node of the National Centre for e-Social Science and its use of a range of JISC-funded services such as national datasets held at Edina or the National Grid Service in developing micro-level simulations of the UK population. The case description presents in detail the social and technical arrangements made within MoSeS, how the project evolved over time and additional activities that develop its research programme into the future. The presentation focuses on the following aspects: background and history of the project, technologies involved, e-learning and training, technological constraints, communication (internal and with stakeholders), community structure and mobilisation, adoption, impact, changes over time, teaching issues, resources and policy input (Barjak *et al.* 2007, pp. 86-95).

4. Fieldwork Pilot

In the eIUS project plan[20] there is a deliberate degree of ambiguity around the intended structure of the project's outputs and the methodology that would be used to develop these outputs. The intention was to reserve judgement until after the conclusion of this scoping study, which would develop and pilot a more detailed methodology.

The project plan did, however, specify an overall framework. This outlined an approach based on the carrying out of a series of interviews and observational studies with active members of the UK researcher community, intertwined with analysis work to produce three different, but related outputs:

- *Experience Reports*, giving concrete examples of the use of existing e-Infrastructure by named individuals or groups of researchers;
- *Use Cases*, based on, and linked back to the experience reports, giving non-technical 'stories' showing how users can interact with e-Infrastructure to achieve specific research goals; and

[19] Accelerating Transition to Virtual Research Organisation in Social Science

20 www.eius.ac.uk/eIUS-ProjectPlan-1a.pdf

- *Domain and Service Usage Models (SUMs),* describing how e-Infrastructure services can be 'joined up' to fulfil specific use cases.

These outputs would be published early and often through a dedicated *community portal*, hosted at NCeSS, to allow for frequent review and validation by project stakeholders.

Note that the treatment of the term *use case* within the eIUS project diverges somewhat from its more typical meaning within a software engineering context where it is taken to refer to a semi-formal technique for describing the functional requirements of a to-be-developed system. In the first place, the eIUS use cases are not describing user interactions with any one system, but with the totality of e-Infrastructure services available to them both within and outside the UK; secondly, the use case is not intended to describe interactions with some future, ideal e-Infrastructure, but with e-Infrastructure as it exists today.

4.1 Initial Discussions

At the e-Uptake and eIUS project kick-off meetings in April and May 2007, it was agreed that the two projects would aim to coordinate fieldwork activities to allow data to be obtained from a wider field of researchers and also to avoid 'over-interviewing' researchers. The projects would work together when identifying researchers, when interviewing them, and when writing up the resulting data as experience reports. The point of divergence would only come after this: the eIUS project using the data to develop use cases and the e-Uptake project focusing on analysing barriers to uptake.

At a fieldwork discussion meeting in July 2007, a more detailed methodology was agreed upon as well as the need to treat it very much as a work in progress and *evolve* it over the course of the pilot study taking into account feedback from informants, reflections of the project team, and experience gained through analysing the resulting data. The interview framework was initially based on a framework developed for the Integrative Biology VRE project (Mascord et al., 2005) which in its early stages was, similarly to eIUS, attempting to elicit current research practice as the basis of efforts to establish a set of high-level requirements for a VRE (albeit for single research domain). This framework was extended by incorporating questions of relevance to the e-Uptake project and introducing a *debriefing* session to enable participants to reflect on the interview itself and its overall organisation.

Methods for drawing a representative sample from the UK research population were also discussed at the July meeting. It was clear that the project, whilst intending to use techniques drawn from social science was not a social science research project in itself. Rather, the project was intending to use these methods in order to reach the very practical outcome of facilitating the increased take-up of e-Infrastructure services. The project team, therefore, came to the consensus that, at least for the eIUS project, less emphasis is needed on achieving a completely representative cross-section of the population (from a social science point of view), and more on uncovering the *successful* and *inspiring* examples that have potential for *transferability* both within and outside the domain they originated. However, it was acknowledged that a certain degree of sampling has to take place in order to ensure a credible spread of disciplines, activities and roles.

4.2 Sampling Approaches

There are two conflicting concerns that need to be met in order to draw an appropriate sample from the UK research population. On the one hand, there is the need to achieve a *large enough* sample through convincing researchers to be interviewed or observed. On the other, there is the need to achieve a *relevant* sample, that is, identify those researchers who are genuinely using e-Infrastructure rather than more generally applicable ICTs. The original eIUS proposal[21] stressed the extensive local evidence base built up through Oxford and Manchester's long-standing involvement in numerous e-Infrastructure services and e-Research projects cutting across the disciplinary spectrum. However, the project plan, written after subsequent reflection, is more candid in describing the significant challenges that lie ahead in terms of securing the participation of researchers outside these two institutions, particularly in

21 www.eius.ac.uk/eIUS-Proposal.pdf

discipline areas where the consortium is less strongly represented. Early on, it was recognised that in addressing these challenges, there is a risk that the project team might lose sight of engaging those deemed most relevant to the project; it was also clear that the team should be aware of this risk and mitigate against it satisfactorily, to ensure the most effective use of the project's limited resources.

Two key considerations in successfully engaging relevant researchers are the *channels* used, and the *messages* first conveyed to researchers when approached. The principal means highlighted in the project plan, via *known service providers,* has one obvious benefit: the team know, a priori, that informants will be genuinely using e-Infrastructure. In this respect, the eIUS project secured letters of support from nine major UK e-Infrastructure service providers[22] and is meeting with representatives from each in order to develop tailored user engagement strategies. However, the problem with this approach is precisely the fact that the services are already known; the method does not try to uncover the use of less well known services developed by individual institutions or research communities outside well-known funding programmes (for example, JISC, DTI, and UK research councils). Unfortunately, identifying the more niche services is less straightforward.

There are various ways to solicit the involvement of users of e-Infrastructure services. For example, by posting to pre-established announcement lists or even contacting users directly, assuming they have consented to the sharing of their contact details. Both methods clearly have the potential to reach a large proportion of the user community. However, they also suffer from being relatively impersonal and have been reported by those working in UK e-Infrastructure outreach and engagement, to result in consistently poor response rates. Another important consideration is the possibility that the types of researchers who respond to such requests are not necessarily representative of the community, and likely to be the more outgoing types, the 'born networkers'. In discussions with the UK e-Infrastructure service providers, OMII-UK and NGS, it was felt that the group of users hardest to engage are the 'quiet' ones, who in many cases are using e-Infrastructural ICTs to groundbreaking effect. Anecdotal evidence suggests that those belonging to this group are more likely to respond when the request comes from an individual who has provided a level of personalised, one to one support. At a meeting with the to-be-director of OMII-UK in August 2007, Neil Chue Hong reported that a more personal approach had worked well in the OGSA-DAI project[23] where support staff had made requests for participation in similar studies to those posting questions on support lists.

Other methods for identifying usage examples (involving both mainstream and niche services) includes going through university or research institution websites looking for press releases giving examples of the use of novel e-Infrastructure in the support of research. Looking at lists of e-Research projects on research council websites (for example, AHRC, JISC e-Research, and UK e-Science), papers at international conferences (e.g. UK e-Science All Hands), and lists of events held at regional e-Research/e-Science centres, as well as any uploaded posters or presentations, is another potential route. Managers and directors of UK funding programmes and e-Science/e-Research centres have already proved a valuable source of intriguing examples due to their connections and influence in the research community. There is also potentially a lot of scope for making use of organisations such as OMII-UK, who through their Project or Area Liaison (PAL) scheme (akin to Microsoft's MVP programme) have 'eyes and ears' embedded in the research community. However, when uncovering these usage examples, particularly the more advanced or cutting-edge examples, there is the question of unpacking whether the technology concerned is *supported* and used by researchers 'in anger' in any real sense or still in the R&D phase. Often researchers can take a dual role and it is crucial that their role on an e-Research project is made clear: are they acting as a source of requirements, tool developer, or as a genuinely active user?

Contacting researchers directly, by posting to faculty email lists, was another method considered for soliciting involvement. However, this route is not being taken forward because the project team feel it would be too resource intensive if carried out systematically across UK HE, likely to result in a poor response rate, and risks resulting in a large number of informants who are not genuine users of e-

22 UK AG Support Centre, VizNet, AHDS, NGS, OMII-UK, MIMAS, EDINA, NaCTeM, DCC
23 www.ogsadai.org.uk

Infrastructure.

Regardless of the route to usage examples, there is also the intention to carry out *snowball sampling*: asking informants in the interview debriefings who else they know using e-Infrastructure services. This can help extend the network beyond the project team's local institutions, to the wider UK research community. This friend-of-a-friend approach can potentially be facilitated through the use of social or professional networking sites, or customer relationship management (CRM) systems because such systems can keep track of when and by whom researchers have been approached as well as record the professional relationships between individual researchers. However, the project is well aware that the use of commercial CRM systems, typically targeted at a sales audience, may not be appropriate in a higher education context. The project is currently seeking advice from the programme on the use of CRM systems and whether it would be feasible to share CRM data with related initiatives.

4.3 Pilot Interviews

In early August 2007, the project began soliciting the involvement of active researchers based at Oxford and Manchester with a provisional target of ten interviews. As illustrated by Table 3, by the end of August, thirteen interviews had been conducted with researchers working in nine distinct academic fields.

Discipline	**Institution**	**Number of Interviews**
Built Environment	Manchester	4
Applied Econometrics	Manchester	1
Condensed Matter Physics	Oxford	1
Engineering Science	Oxford	1
Corpus Linguistics	Oxford	1
Computational Biochemistry	Oxford	2
Educational Studies	Oxford	1
Evolutionary Biology	Oxford	1
Materials Science	Oxford	1

Table 3 – Discipline Representation in the Pilot

Nine informants (in the built environment, applied econometrics, engineering science, corpus linguistics, and computational biochemistry) were direct colleagues or acquaintances of members of the project team. Two (materials science and condensed matter physics) agreed to participate following an email request forwarded by an Oxford e-Infrastructure service provider. One (evolutionary biology) made himself known at the NGS users' forum in June 2007, where there was a presentation on eIUS and e-Uptake. The educational studies researcher was contacted following a recommendation made by a member of the Oxford University Computing Services (OUCS) working on requirements gathering for future OUCS services.

In the interview debriefings, the project team asked informants what the main reason was behind their decision to participate. At the beginning of the project, the project team felt that only by offering some incentive would researchers offer to participate, for example, publicity, networking opportunities, or having a constructive impact on the service providers. As illustrated by table 4, however, what the project team did not expect was the importance of *goodwill* as a motivating factor. Four cited this as a reason they had decided to participate, with two saying that 'giving something back' was the main reason they had decided to take part. Admittedly, the majority of informants in the pilot study were acquaintances of members of the project team, and the latter two had received a considerable amount of one to one support from an individual running an Oxford e-Infrastructure service, seeing eIUS as an extension of its outreach and engagement activities. Another factor not initially considered relevant was

the opportunities the project might bring for sharing and reflecting on the research methodologies that might help to address a perceived inconsistency in the level of research methods training across UK higher education.

Motivating Factor	Total
Networking with other researchers.	6
Reflecting on, or sharing research methodologies.	5
Goodwill - giving something back, helping each other out, returning a favour.	4
Publicity.	4
Help improve the services by highlighting problems/issues.	3
Representing a specific king of need – e.g. particular research community, particular kind of usage need.	2
Empathy with SP point of view – understanding of need for sustainability.	1

Table 4 – Benefits Cited

The first five interviews were arranged over the phone and scheduled to take place in Manchester in early August 2007 over two days, with two hours reserved for each interview. Typically, three people (excluding the interviewee) would be present at each interview: one person asking questions for the eIUS project, a second asking questions related to the e-Uptake project, and a third taking notes. To supplement the paper notes, a MiniDisc recorder was used to produce an audio record of the interview. As specified in the original interview framework, each interview began with a short introduction where the overall aims of the project were explained and an overview of the interview questions given. Before the interview itself, and before the recording equipment was activated, participants would then be given the consent form and time to read and sign it. The eIUS phase of the interview followed a 'day in the life' format, focusing on the research narrative but drilling down into specific tasks, and the particular tools and technologies used to perform these tasks. A basic categorisation of services allowed the interviewer to keep a mental note of major service types covered and ask briefly at the end about any major types not already covered by the research narrative. In the e-Uptake phase of the interview, researchers were questioned further about any barriers or issues mentioned, including those relating to the technologies and to the carrying out of their research in general.

As mentioned earlier, each interview concluded with a debriefing where participants were asked specific questions about the interview itself. Feedback collected as part of these sessions related both to the structure of the interview itself as well as to its overall organisation, and was collated incrementally in an informal *lessons learned* document. In terms of the organisation of the interviews, a recurring request was for more material to be sent in advance. Several participants stated that they would have preferred something in writing and to be sent a basic overview including a list of questions. Several preferred the idea of being sent at least the consent form in order to allow more time to read it at their leisure rather than under the pressure of the interview. In terms of the structure of the interview itself, there was a feeling that the 'day in the life' format was not particularly suited for those working across multiple projects, and even for those working on a single project, such as a PhD, there was no typical day because it depended very much on the phase of the research. It was recommended that rather than starting with the 'day in the life' question and moving on to how this relates to the overall research lifecycle, it would be better to arrange it the other way round, with questions on day-to-day research activities positioned explicitly within each stage of this research lifecycle. Either way, it was agreed that the research narrative should be the starting point and questions about tools and technologies made within this context rather than starting with the tools and then asking about the research. Two further comments related to the duration and location of the interview. It was suggested interviews should be no longer than an hour, and that it may be helpful to carry them out *in situ* to see the context the researcher is operating within. The former suggestion clearly has major implications regarding the level

of detail attainable within the interview itself, although many appeared happy to respond to follow-up requests for more information.

For the final eight interviews, a new interview framework was developed on the basis of this feedback. This began by asking the informant to give an overview of their research area and examples of specific research questions their research addresses. The next phase of the interview involved going through each stage of the research lifecycle, asking the participant to give examples of the kinds of day-to-day research tasks performed and supporting tools and technologies used in each stage. The only change to this framework, following interview six, was to ask the participant to state clearly which particular project the particular research task referred to, in order to avoid confusion. There was no further feedback on the interview methodology or changes to the framework between interviews 7 and 13.

4.4 Developing Experience Reports and Use Cases

In early September 2007, a data session was held to discuss the pilot fieldwork data collected in August. This meeting had two goals: one was to progress the project by agreeing on a working format for the use cases, the other was to decide on content for a paper accepted by the International e-Social Science Conference 2007. In advance of this meeting, notes from the interviews were typed up and two prototype use cases developed. Time constraints meant that the use cases were constructed directly from the interview notes rather than following the writing of experience reports as was originally intended. The majority of the meeting was spent going through the two use cases in order to understand how they were created from the evidence with the hope of making it a more repeatable process. As no experience reports had been prepared in advance, a format for these was not agreed upon at the meeting. It is likely, however, that the experience reports will be structured around the generic research lifecycle that has proved to be a successful way of structuring the evidence gathering activities themselves.

The original intent of the experience reports developed in the proposal and project plan was to serve as multimedia written-up versions of the raw data obtained from the interviews and observational studies. The use cases were intended as 'stories' *based* on this evidence. However, for the most part, informants did not tell stories in their interviews, instead they gave examples of the *types* of research tasks performed, *tools* used, and the *decisions* that have to be made in the context of overarching research questions. It became apparent therefore, that it was not going to be feasible to *distill* realistic use cases from the data alone without further research into what is typical within the field, and even after that, there would inevitably be a degree of invention in the resulting accounts. To ensure the resulting use cases would be considered *believable* by members of the research community, it was clear that they must be *validated* with the original informants. Changes suggested would be made and any additional contextual information integrated back into the experience reports. Links from fragments of the use cases would be linked with corresponding fragments in the experience reports to make explicit where particular statements are actually grounded in the data collected.

Appendix A gives four examples of use cases generated from the interview data. The first example, from the field of engineering science, was also the first example discussed in the eIUS data session. It was agreed that its length was about right, not over-long, and the numbered steps made it easy to parse. The first step, which sets the story in motion, gives an initiating event, names the main actors, their respective research interests, and situates the story in time and space by giving a specific month, time of day, and location. The remaining steps develop the story from this initial situation. The process of creating the use case was relatively straightforward for this example because the informant had provided a good level of insight into how research projects in this domain are initiated and develop. The only major change suggested by the informant was for the actors (in step 4) to have a meeting, where the clinician explains what the images represent, and what she hopes to find in them. The second use case discussed at the eIUS data session, from the field of education studies, was felt by those present to be more akin to an experience report, or an introduction to an experience report, in that it described the overall research agenda rather than telling a specific story. It was therefore agreed that the first example be taken forward as a working prototype for the use cases in the eIUS project.

When writing the use cases, the project team had to make a series of editorial decisions on whether or not to include the names of real institutions, research questions, journals, and ICTs. Each use case presented in Appendix A deliberately adopts a slightly different editorial policy in order to show the implications of making choices either way. In terms of the ICTs featured, the project team agreed that it is important to acknowledge that the use case is set more or less in the present day and as such should recognise non-technological or mundane solutions to problems where these have been cited in the interview data rather than always giving a more technologically advanced example. For example, in the image analysis example, a DVD was sent through the post rather than data being shared using Secure FTP or the Storage Resource Broker (SRB)[24]. Other editorial considerations relate to the level of detail to include in the use case. On validating the use cases, informants have frequently come back with a wealth of additional contextual information much of which belongs, and should be integrated back into, the experience reports. Links to this information from the use case, will allow the consumer to continue reading at particular points, if they so wish. As the project team is not yet committing itself to a fixed set of editorial guidelines, it was agreed that all decisions made in the development of the use cases be recorded to enable a continuous review to be maintained on the editorial process throughout the project.

Another key consideration discussed was how use cases would be marked up, behind the scenes, in order to make them discoverable, and to move towards a more dynamic form of use case alluded to in the eIUS proposal. For example, it was agreed that all instances of *research activities* and *technologies* should be linked to controlled vocabularies. The external appearance of this mark-up would take the form of links to human readable glosses designed for a non-specialist audience. *Decision points* associated with choice of technology would also be annotated to show *why* one tool was chosen over another.

4.5 Developing Service Usage Models

Two workshops on aspects of the e-Framework were attended by project members: a general informational workshop in Birmingham on 31 May 2007, and another focusing more specifically on Service Usage Models (SUMs) in London on 4 June 2007. Development of SUMS, however, was not explored in the scoping study. It was felt premature for this to be part of the study, when the project had not yet decided on a format for the experience reports and use cases, designed to be the basis of the SUMs. There was also the feeling that the project should wait until the community had reached more of a consensus on best practice regarding the structure and content of Service Usage Models.[25]

4.6 Treatment of Audio Data

Although it was always the intention to produce audio recordings of the interviews carried out with researchers, the budget set in the proposal did not have an allocation that would enable the audio recordings to be professionally transcribed. The audio recordings were intended only as supplementary to the notes, to be used for example, to help fill in any gaps, resolve discrepancies, and pick out quotes. However, when members of the project team attempted to go back to the audio when writing a use case for the e-Social Science 2007 paper, this process proved much more time-consuming than was originally anticipated. Revealingly, however, the result of this process was to include direct quotes from interviews in call-out boxes, which demonstrates the value and impact of citing the actual words used by informants in interviews.

It is currently unclear whether there is any real longer-term benefit to be gained from transcribing all interviews. However, there is a consensus that if the audio recordings are going to be usable in future, they would need to be transcribed, ideally by external contractors so the project team is not distracted from the critical task of analysing the resulting data. Transcribing the audio data also has the potential to make the data more *reusable* because information that might not have been originally considered relevant is captured. Although it was recognised that carrying this out on a larger scale would need

24 www.sdsc.edu/srb/index.php/Main_Page

25 The e-Framework for Education and Research has since made available a small number of draft SUMs relating to the research domain (https://e-framework.usq.edu.au/users/wiki/DevelopmentSUMs).

additional funding, and formal agreement from JISC, it was agreed that the method should at least be evaluated with a maximum of two interviews. If this proved successful, the project would retrospectively make a case and seek funding for its use on a larger scale.

The possibility of making greater use of qualitative data analysis software such as NVivo and Atlas.ti to manage interview data was also explored. At the time of writing the team believes the cost of learning how to use these tools will begin to pay off once the number of interviews being analysed increases. Consequently, the project now intends to evaluate these tools in the next phases of the project.

4.7 Capturing Video Data

Use of video in the next phase of the project is expected to make the experience reports more compelling by visibly showing how researchers actually interact with e-Infrastructure, thus overcoming the somewhat *operationalised* nature of the accounts. The project team recognises, however, that this technique will not always be appropriate and will clearly be unsuitable for illustrating activities that are fundamentally non-visual in nature, or made up of a series of isolated actions performed over an extended timescale. The technique is expected to be most valuable for illustrating particular segments, such as setting up and running a simulation, or performing a particular kind of textual analysis on a corpus, for example. During such video vignettes, those under observation are expected to *talk aloud*, and give a commentary on the tasks they are carrying out for the benefit of those viewing the video.

4.8 Data Publication

The intended primary means of publishing the use cases, experience reports, and SUMs, as described in the proposal and project plan was through a community portal, to be hosted at NCeSS. Although the establishment of this portal was originally intended to take place in parallel to the scoping study, it became clear that the detailed requirements for this portal could not be identified until a consensus had been reached on a working format for the use cases and experience reports, and a reasonably large corpus of these items, established. The use cases have already proved to be particularly effective at provoking discussion on the typical use of e-Infrastructure by informants and it is hoped that the format for these use cases can be further refined through the organisation of workshops designed specifically to engage the wider research and research-support communities.

Ethical considerations made explicit in the participant consent forms, demand that any non-anonymised data is sent to the individual concerned for approval before publication. The project team also intends to offer an acknowledgment to informants who have assisted in the construction of the use cases. Another consideration of the community portal not explored in the scoping study is how to integrate third party data, such as use cases produced by other initiatives.

5 Conclusions and Next Steps

This report has described how a methodology for discovering and publicising the use of advanced ICT by UK researchers was piloted and further refined through a small-scale evidence-gathering pilot at Oxford and Manchester. Crucially, the report helps to scope and prioritise the efforts of the project, by giving a set of criteria for, and examples of ICTs that are within, as well as out of scope. A working definition and format for the eIUS use cases is given, as is four prototypes drawn from the data collected during the pilot. Also noted, is the intention to keep the overall methodology, including the editorial policy adopted in the creation of the use cases, under continuous review to enable it to adapt to the challenges of a UK-wide roll out.

In the next phases of the project, the use cases will be published through a dedicated community portal hosted at NCeSS and the corpus further reviewed at series of community engagement events to be organised in conjunction with the e-Uptake project. The community portal will be designed to allow consumers a more meaningful interaction with the use cases, including the ability to trace back to the evidence, pull up human readable descriptions of the technologies and research activities described, and

where appropriate, view video vignettes of researchers actually interacting with e-Infrastructure.

6 Acknowledgements

eIUS, which is funded by JISC under their e-Infrastructure Programme, is led by the Oxford University Computing Services (OUCS) in conjunction with the Oxford e-Research Centre (OeRC), and in partnership with the National Centre for e-Social Science (NCeSS).

The project would like to acknowledge the help of the following researchers in compiling the example use cases presented in Appendix A: Dr Simon Peters, Dr Ylva Berglund, Dr Vicente Grau, Dr Kaihsu Tai, and Dr Kia Balali-Mood.

The questions on the research lifecycle incorporated in the interview framework were based on a research lifecycle presented by Matthew Dovey at the 2nd JISC VRE Programme Meeting, Cambridge, 6 July 2005.

Finally the project would like to acknowledge the assistance of the e-Uptake project, led by NCeSS in partnership with NeSC and AHeSSC, for their help and cooperation in the joint fieldwork carried out to date.

7 References

Barjak, F. *et al.* (2007). *Accelerating Transition to Virtual Research Organisation in Social Science (AVROSS) – M4 Final Report*. Draft v.2.2 of 27.10.2007. EU Service Contract No. 30-CE-0066163/00-39. http://plattformen.fhnw.ch/avross

Fleck, J. (1988). *Innofusion or diffusation? The nature of technological development in robotics*. Edinburgh Programme on Information and Communication Technologies (PICT) Working Paper No. 7, Edinburgh University.

Gamma, E., Helm, R., Johnson, R. and Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading, MA.

Hughes J.A., Randall D., and Shapiro D.(1992). Faltering from ethnography to Design. *In Proceedings of CSCW'92*, ACM Press, pp. 115-122.

Hughes, J., O'Brien, J.,Rodden T, Rouncefield. M and Blythin, S (1997). Designing with Ethnography: A presentation Framework for Design. *In Proceedings of Designing Interactive Systems '97*. ACM Press Amsterdam. pp 147-159

Kalawsky R.S., Holmes I.R., O'Brien J., Lewin M., Armitage S., Goonetilleke T.S., Nee S.P. (2006). *Human Factors Audit of Selected e-Science Projects*. e-Science HF Audit Report. Available at: http://syseng.lboro.ac.uk/jisc/humanfactors/

Mascord, M., Jirotka, M., Sieunarine, C. (2005) *Integrative Biology VRE Initial Analysis Report*. Available at http://www.vre.ox.ac.uk/ibvre/IBVRE%20Initial%20Analysis%20Report.pdf

Martin, D., Rodden, T., Rouncefield, M., Sommerville, I. and Viller, S. (2001) Finding patterns in the fieldwork. *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work*, pp. 39--58, Bonn, Germany: Kluwer Academic Publishers.

Martin, D. and Sommerville, I. (2004). Patterns of Cooperative Interaction: Linking Ethnomethodology and Design. *ACM Transactions on Computer-Human Interaction* 11(1), pp. 59-89.

Newhouse, S., Schopf, J.M., Richards, A. and Atkinson, M. (2007). *Study of User Priorities for e-Infrastructure for e-Research (SUPER)*. In: UK e-Science Technical Report Series (ISSN 1751-5971).

Available at www.nesc.ac.uk/technical_papers/UKeS-2007-01.pdf

Plowman, L., Rogers, Y. and Ramage, M. (1995). What Are Workplace Studies For? In Marmolin, H., Sundblad, Y. and Schmidt, K. (eds.) *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work*. September 10-14, Stockholm, Sweden.
Available at: http://www.ecscw.org/1995/20.pdf

Sørensen, K. H. (1996). *Learning technology, constructing culture: socio-technical change as social learning*. STS working paper 18/96, University of Trondheim: Centre for Technology and Society.

Stewart, J. and Williams, R. (2005). The Wrong Trousers? Beyond the Design Fallacy: Social Learning and the User. In Rohracher, H. (ed.) *User involvement in innovation processes. Strategies and limitations from a socio-technical perspective*. Profil-Verlag, Munich.

Trigg, R., Blomberg, J. and Suchman, L. (1999). Moving document collections online: The evolution of a shared repository. *Proceedings of the 6th European Conference on Computer-Supported Cooperative Work*. pp. 331-350.

van Welie, M. and van der Veer, G. C. (2003). Pattern Languages in Interaction Design: Structure and Organization. *Proceedings of Interact'03*. September 1–5, Zurich, Switzerland. Amsterdam, The Netherlands: IOS Press.

Viller, S. and Sommerville, I. (2000). Ethnographically Informed Analysis for Software Engineers. *International Journal of Human-Computer Studies*, ISSN: 1071-5819, Volume 53, pp.169-196.

Voß, A., Procter, R. and Williams, R. (2000). Innovation in Use: Interleaving day-to-day operation and systems development. *Proceedings of the Participatory Design Conference*. New York, Computer Professionals for Social Responsibility.

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

8 Appendices

Appendix A – Prototype Use Cases

In these examples, underlined elements are based either on the original interview data or on feedback sent by the informants in response to draft versions of the use cases. Key *activity types* are highlighted in *italics* and **relevant ICTs** in **bold**.

Use Case 1 – Engineering Science

Narrative

1. Alex, a researcher working in medical image analysis is phoned up one evening in June by his friend Alison, who the day before *produced a sequence of MRI brain images* from hospital volunteers of an unprecedented high resolution.
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*, 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.
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** .
7. Alex *downloads this code* from the **Insight Toolkit (ITK),** 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.

> "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."

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 *parallelise the code*, that is, carry out further work to have different compute nodes in the cluster analyse separate sections of the images.
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.

Relevant ICTs

ICT	Comments
Insight Journal	An open access image analysis journal with associated code repository.
Insight Toolkit (ITK) [26]	An open-source software toolkit for performing registration and segmentation. Code is sourced from submissions to the Insight Journal.
Local Cluster	A high performance compute resource provided locally.
MatLab	A numerical computing environment and programming language.
BrainWeb [27]	A simulated brain database. Not included in use case as reported to be used only occasionally.
Field II [28]	An ultrasound image simulator. Not included in use case as reported to be used only occasionally.
Storage Resource Broker (SRB)	SRB is a technology for managing collections of files distributed across multiple organisations and heterogeneous storage systems. In the use case step 4, Alison could have used Storage Resource Broker (SRB) or Secure FTP to share the images. However, she chose not to, preferring to rely on a DVD sent in the post. Possible reasons for this might include: - Lack of awareness of SRB or Secure FTP. - Lack of access to a supported SRB installation. - Institutional firewall blocks the network ports needed for SRB or Secure FTP. - Lack of urgency.

Commentary

In this first example, the research question is loosely based on one of the examples given in the interview data also relating to analysing blood flow, but in the heart. The research methodology, characterised by the analysis of sequences of high-resolution MRI images from hospital volunteers, has been kept exactly as it was described in the interview.

Many other elements of the use case are taken directly from the interview data. These include:

- the means of sharing the images (by sending a DVD in the post);
- the suggestion that test images are not always available within ITK (due to patient confidentiality issues);
- that the analysis of the larger images is not possible on his workstation due to a lack of memory;
- the need to parallelise the code to make it run on the local cluster;
- the use of the Insight Toolkit and Insight Journal for accessing and publishing image analysis code.

Comments by Informant

The informant said that overall the use case was 'quite accurate'. He did, however, have a number of comments and suggested changes to the use case that were subsequently incorporated. These included:

26 www.itk.org
27 www.bic.mni.mcgill.ca/brainweb
28 server.oersted.dtu.dk/personal/jaj/field

- the detail about the two collaborators first arranging a meeting to discuss the images;
- the detail about the contrast in the images decreasing to the point where existing analysis algorithms fail (high resolution in itself is not normally an issue);
- the detail about the analysed images being validated against an independent manual analysis.

Other Editorial Considerations

Element	**Usage**
Links to direct quotes?	Yes
Year?	No
Month?	Yes
Time of day?	Yes
Location given?	No
Real institutions named?	No
Real journals named?	Yes
Real conferences named?	No

Example Use Case 2 (Applied Econometrics)

Narrative

1. Sally, a labour economist, based at Stirling University and with a speciality in the welfare of ethnic minorities is waiting patiently. Three years ago at a conference she met John, an econometrician and they decided in passing to look each other up the next time an opportunity for a new research project arose. Sally is hoping that together they can bring mutually complementary perspectives to each other's research.
2. Arriving early at Southampton University for a face-to-face meeting with John to discuss the writing of a new proposal for their first prospective project together, Sally decides to have a coffee and *check her email* using the wireless network in the university canteen.
3. For the past two weeks, Sally has been *doing some modelling* using an **innovative web portal** developed by John's team. This web interface allows Sally to *perform a kind of statistical modelling* previously only accessible to those with an extended knowledge of the Linux command line, high performance computing, and Fortran. (Sally has never used the command line.)
4. Sally *launches the portal* in order to check over the data she wanted to discuss with John. Because Sally had already signed onto her university's bibliographic and data access system, she does not need to register with the portal or the underlying computing systems it uses because it is also controlled under the UK access management federation for higher education.
5. *Visualising the study* she had prepared earlier, the results, presented over a map of the UK show the distribution of inequality that appears to validate her own research. However, the model prepared by John does not incorporate a particular macro-economic variable of central importance in her research.
6. John arrives on time and they decide, given the quiet conditions in the canteen, and the fact that Sally already has the portal open on her laptop, to have their meeting there and then, rather than going up to John's office.
7. They *discuss the data* and agree that the existing portal does not incorporate a number of key regional macro-economic variables that would control for regional price differentials (for example, the fact that London is generally more expensive than Scotland). John explains that addressing this would involve incorporating an additional dataset but agrees to *make the changes* necessary within the next couple of weeks.
8. Over the next two months, the portal proves invaluable for Sally for around 80% of the research questions she wants to ask. However, for the remaining 20%, Sally has to ask John to *add bespoke improvements and analysis options* for the model she was originally using. Nevertheless, the portal removes a major bottleneck and allows Sally to make considerable progress in the majority of cases, without any intervention from John.
9. In May they have together collated sufficient data and are ready to *write the results up* in their first

academic paper. Sally *selects the subset of the archived cases* that is most pertinent for their study of inequality. This is *presented* at a series of UK and European economics conferences and the first article appears in the Scottish Journal of Political Economy.

Relevant ICTs

ICT	Comments
Innovative web portal for labour economics	The GEMEDA[29] demonstrator portal, originally developed under the NCeSS pilot demonstrators project, closely resembles the system described in the use case with the exception that GEMEDA is not controlled under the UK access management federation. GEMEDA requires users to possess three independent security credentials: (1) a GEMEDA service username and pass-phrase, (2) a valid UK e-Science Certificate[30] to access the National Grid Service underpinning the service, and (3) an Athens[31] account to access the datasets exposed by the GEMEDA service. Although the GEMEDA service at the time of writing is still publicly available, lack of funding means it is no longer properly supported as such.
High performance computing (HPC) resource underpinning the portal	The name of an actual HPC resource is not given because the researcher interacting with the portal should not need to know which particular HPC resource is used. As stated above, the HPC resource underpinning the GEMEDA portal is the National Grid Service (NGS).
UK Access Management Federation for Education and Research[32]	The federation provides a single solution to access online resources and services for education and research.

Commentary

The research question in this use case was taken directly from the interview data, as was the nature of the collaboration between the applied econometrician and the labour economist. The use case hints that the portal is more of a prototype in nature than a properly supported service. However, the use case is slightly idealistic in the sense that the portal is integrated with the UK Access Management Federation, which at the time of writing, is still in its early stages of becoming established across UK higher education.

Comments by Informant

In this example, the informant who had contributed to the development of the GEMEDA portal, made several changes aimed at making it more believable. He also gave additional contextual information that is included below. The informant suggested GEMEDA not be named within the use

29 pascal.mvc.mcc.ac.uk:9080/gemeda
30 www.grid-support.ac.uk/content/view/23/182/
31 www.athens.ac.uk
32 www.ukfederation.org.uk

case itself while other elements of the use case were for illustration only.

The usual mode of working for the labour economist would be to use a standard proprietary statistical package (for example, Stata[33], SPSS[34], SAS[35]), or a package add-in to perform the modelling required. The labour economist would usually work with secondary data supplied by providers such as the ONS[36] (Office of National Statistics) or the ESRC Data Archive[37] at Essex. She would need to clean the data herself, maintain working copies and manage her modelling and results through package scripts or records of interactive sessions. The labour economist would have a better understanding of the data she works with than the applied econometrician, and the literature and modelling fashions relating to the substantive topic under study. The modelling and the production of the labour economist's results are, however, restricted to that available within her package of choice. The applied econometrician, by contrast, is familiar with the package the labour economist uses and can naively extend it by using simple scripts or macros. However, he can also write his own bespoke modelling code using either specialist matrix programming languages (for example, Gauss[38], Ox[39], Matlab, Octave[40]), statistical languages (for example, R[41]) or third generation programming languages (for example, Fortran, C). He may also be aware of weakness in the algorithms used in the proprietary packages and knows how to parallelise serial code.

Commenting on step 9, the informant noted that the portal is self-curating and annotating so that details of all the cases Sally has investigated are kept along with the results of the requested analyses. He also made the point that Stirling University does research in applied labour and other microeconomics, one of its professors used to be on the board of the Journal of Political Economy.

Other Editorial Considerations

Element	**Usage**
Links to direct quotes?	No
Year?	No
Month?	No
Time of day?	No
Location given?	Yes
Real institutions named?	Yes
Real journals named?	Yes
Real conferences named?	No

Example Use Case 3 (Computational Biochemistry)

Narrative

1. Martin, a computational biochemist based at Glasgow University is *browsing* the **Protein Data Bank** one morning in December when he comes across a new entry from a colleague based on the East Coast of the USA.
2. Martin knew his X-ray crystallographer friend Thomas was working on determining the cartesian coordinates for this particular protein (human lung cytochrome) but did not realise he was ready to upload the coordinates to the **Protein Data Bank**. He had been waiting for these coordinates because he wanted to run a simulation to *validate some experimental results* recently obtained by another

33 www.stata.com
34 www.spss.com
35 www.sas.com
36 www.statistics.gov.uk
37 www.data-archive.ac.uk
38 www.aptech.com
39 www.doornik.com/ox
40 www.octave.org
41 www.r-project.org

collaborator, Eric, based in the Netherlands.

3. Martin *phones up Thomas* to get the full story, which involves Thomas achieving an unexpected breakthrough in obtaining a viable sample of cytochrome, expected to take many months longer. Thomas did intend to let Martin know earlier but they needed to get the results submitted in a hurry to be included in their research institution's quinquennial review, which has major funding implications.
4. Three years ago, Martin had *performed a similar simulation* of mouse hepatic (liver) cytochrome, using the NAMD molecular dynamics simulation application but no longer had the relevant scripts and input configuration files to hand. Fortunately, Martin was an early depositor to the simulation archive **BioSimGrid**, a fully curated repository of molecular dynamic simulations.
5. Martin *searches* ***BioSimGrid*** for the keyword, 'cytochrome', and finds that his three-year-old simulation is the second entry in the results. He *clicks on 'download'*, and *saves the simulation system* 'mhC-20031003.pdb' along with the accompanying trajectory 'mhC-20031003.xtc' (in the format of the Gromacs, not the original NAMD, molecular dynamics simulation application) to his desktop.
6. By *clicking on this file*, **the molecular viewer, VMD**, is opened immediately. Martin *takes a look* at the protein and the surrounding solvent (water) molecules and *replays the simulation*. The visualisation tallies with what he remembered from the paper he wrote 3 years ago. However, he needs to *replace the mouse hepatic protein with the human lung variant*.
7. Martin then *superimposes the original mouse protein with the human variant*, plugging in Thomas's coordinates from the **Protein Data Bank**, and *submits the new simulation* to his lab's **Local Cluster**.
8. Four hours later, Martin receives a new email with a link to the new simulation set. He *downloads it and replays it*. Unfortunately, the result does not tally with Eric's experimental results.
9. Phoning up Eric, it turns out that the most likely reason for this is because reason for this is because the protonation states for some of the residues in the protein is not simulated realistically in accordance with the physiological pH value. Making the simulation more realistic means further work to the input configuration of the simulation.
10. Eric and Martin work over the next two months to *validate Eric's experimental results* without success. Eventually Eric agrees to *re-do his original experiments*, under more constrained and therefore known conditions.
11. The monitoring of an additional variable through the experiment, leads to the need to alter the simulation to match this, and Martin has to *make a series of modifications to the protein* through the molecular viewer. In particular, the protonation states of some of the residues in the protein and the solvation conditions by water molecules needed to be carefully determined and adjusted, using both existing data in the literature (in the PubMed literature database) and Eric's new results.
12. After two months of work, Martin finally obtains simulation results that tally with experiment and they decide to *publish* a conference paper on the findings.
13. Martin uses **VMD** and its rendering plugins to *create a ray-traced animation* of the results, which is embedded in a presentation he gives to an international biophysics conference in Long Beach. On receiving favourable feedback, Martin and Eric spend the next three months *extending their study* taking into account suggestions.
14. The full results are eventually *written up* within a well-respected computational biochemistry journal.

Relevant ICTs

ICT	Comments
BioSimGrid[42]	A fully curated repository for molecular dynamics simulations (trajectories).
Protein Data Bank[43]	A fully curated repository for the processing and distribution of 3-D structure data of large molecules of proteins and nucleic acids.
Local Cluster	A high performance compute resource provided locally.
VMD	A molecular visualisation program for

42 www.biosimgrid.org
43 www.pdb.org

	displaying, animating, and analyzing large biomolecular systems using 3-D graphics. VMD is considered within the scope of eIUS because it has the capability to work with a molecular dynamics program running on another computer in order to display the results of a simulation as they are calculated. However, it is not possible to submit new simulations directly from VMD. New simulations must be submitted manually, demanding a good appreciation of the command line, cluster computing and scripting languages.

Commentary

This use case is an example that illustrates the way in which researchers can interact and combine e-Infrastructure services provided at different levels, from infrastructure provided locally (in the case of local clusters) to infrastructure hosted internationally (in the case of the Protein Data Bank). All the ICTs featured in this use case were given as examples by the two informants based within a computational biochemistry lab at Oxford. Local clusters are highlighted because the informants were both based within a well-funded lab that had three powerful in-house clusters, in addition to a large HPCx[44] allocation. This meant they did not see the benefit of going through the procedures required to gain access to the National Grid Service.

The overall shape of the research described by the use case - comparing experimental results with simulation studies - was taken from the interview data. However, a lack of concrete examples of typical proteins meant an appropriate example was chosen from the computational biochemistry literature.

Comments by Informant

The basic structure and flow of the use case remained unchanged following review. However, the original use case had a series of gaps that were filled in by the informant. These details included,

- in step 4, the incorporation of mouse hepatic (liver) cytochrome as the protein featured in the previous simulation;
- in step 5, the details of the simulation set downloaded from BioSimGrid;
- in step 9, the reason the simulation did not tally with experiment;
- in step 11, the details of how the simulation was altered to match the experiment.

Other Editorial Considerations

Element	**Usage**
Links to direct quotes?	No
Year?	No
Month?	Yes
Time of day?	Yes
Location given?	No
Real institutions named?	Yes
Real journals named?	No
Real conferences named?	No

Example Use Case 4 (Corpus Linguistics)

44 www.hpcx.ac.uk

Narrative

1. Ella, a corpus linguist based at a large research-led university, with a specialty in modal verbs, receives an email in the early hours from a corpus linguist based in Japan who she has never met before.
2. Karl, who has a particular interest in academic writing wonders whether Ella might be interested in collaborating with him in order to do a comparison of the expression of modality in academic writing compared with British English as a whole. He had just read an article she has just published about modal verbs in school essays. He found it through Google Scholar.
3. The following day, Ella replies to Karl saying she is, in principal, open to such a collaboration, although she makes the point that with limited funds, they are unlikely to be able to meet each other in person in the near future. However, they are not unused to this style of informal collaboration. (Research for both Karl and Ella is an unfunded, out-of-hours activity).
4. They plan to use a **web-based version of the British National Corpus, hosted at Brigham Young University (BYU-BNC)**, which Karl already has some experience with. Ella, on the other hand is more familiar with **BNCWeb**, an alternative web-based interface to the BNC available to members of her university.
5. The next day, Ella visits the **BYU-BNC site**, and phones up Karl so that he can walk her through it. Karl explains the basic search options, and how the search results can be examined in different ways. He also describes how to use some of the more advanced features.
6. Over the next six months, they divide up the analysis work between them. Ella focuses on analyzing modal verbs (which she has expertise in), whilst Karl works on analyzing other means of expressing modality.
7. For the first month, they analyse and swap data via spreadsheet files sent over email. However, on a recommendation from a colleague of Ella's they trial Google Documents, which incorporates a web-based spreadsheet editing package and allows them to manage versions more efficiently.
8. They decide to present their preliminary findings at a well-known Corpus Linguistics conference. In their talk, some audience members ask questions about whether they have thought about considering whether their findings also hold true in American English, which could shed light on whether academic writing is closer to American or British English.
9. Fortunately, they find out that the same person who developed the **BYU-BNC** is also in the process of developing a comparable corpus of American English. As they are so happy with the way the previous study worked, they contact the developer of the new corpus and he agrees to allow Karl and Ella to beta-test the new system before its official release.
10. After only two months of work using this new corpus, they already have some interesting results. However, the results have to be treated as tentative because the corpus is still only partially constructed. Once it is released, they intend to validate their results with the full corpus and publish their findings in the International Journal of Corpus Linguistics.

Relevant ICTs

ICT	**Comments**
Mark Davies / Brigham Young University Interface to the British National Corpus (BNC)[45]	A publicly accessible web interface to the British National Corpus hosted at Brigham Young University.
BNCWeb[46]	An alternative web interface to the BNC. Equivalent installations of the software (to be released under the GNU GPL[47]) exist at various UK institutions. However, due to BNC licence restrictions, each installation is accessible only to members of the institution hosting the installation. Those with no local installation

45 corpus.byu.edu/bnc
46 www.bncweb.info
47 www.gnu.org/copyleft/gpl.html

	would need the specialist skills required to install their own copy. The BNCweb interface is considered to be more advanced than the BYU interface. For example, the BNCweb interface allows sorting on terms up to five words away from the search term, can produce ranked lists of collocates, and allows export of data for further analysis offline. However, the BYU interface is publicly available and very fast.
WordNet[48]	The BYU interface to the BNC integrates with WordNet, a semantic analysis of words. For example, it allows searches on all words than mean 'big'.

Commentary

Much of this use case changed following a review with the original informant who suggested alternative research questions, more appropriate corpora, and more likely (physical) locations for the researchers. What remained the same, however, was the idea of illustrating a geographically distributed collaboration arising between two people who had never met. This feature of the use case was taken from an actual example given by the informant. Also taken directly from the interview data was the description of the research as an unfunded, out-of-hours activity.

There was some debate at the eIUS data session about whether the use of concordance tools represented genuine use of e-Infrastructure. This was because use of BNCweb, considered as e-Infrastructure because it is web-based, did not come out very clearly in the interview data. However, in a follow-up email, the informant confirmed that she 'would not have been able to do her research without BNCweb'. For reasons explained below the use of BNCweb was replaced with use of the similar Brigham Young University interface in the final version of the use case.

The use of Google Documents was included in this example because the informant had indicated their genuine intention to use this tool to improve the management of result data held in spreadsheet files.

Comments by Informant

In this example, the research question, ICTs used, and locations were all changed following review to make the use case more believable from a corpus linguistics point of view.

In the first version, the research question, constructed from evidence gathered through desk research rather than from the interview data, concerned the differences between newspaper prose between Manchester and New York in the 1930s. The informant suggested that this was unrealistic because firstly, the BNC was not designed to be representative of UK newspaper journalism but of British English (1970-1995) as a whole and secondly, there is no known corpus of 1930s New York newspaper journalism. As an alternative, the informant suggested that the BNC corpus be used to compare expressions of modality in academic writing compared with British English as a whole. Originally, the BNCweb interface was featured. However, this was replaced with the BYU interface because it is freely available.

The location for Karl was changed from New York because the informant doubted whether a linguist based in New York would be interested in British English and suggested Europe or Japan might be a more likely location, there being several researchers studying British English in these regions. Japan was chosen to highlight the geographically distributed nature of the collaboration.

48 wordnet.princeton.edu

The use case was also extended by adding the idea of extending the study by carrying out a similar analysis on an American English corpus written by the developer of the BNC interface. However, it is an open question whether including fictional interactions with real people is appropriate, even where such people are referenced only indirectly.

Other Editorial Considerations

Element	**Usage**
Links to direct quotes?	No
Year?	No
Month?	No
Time of day?	Yes
Location given?	No
Real institutions named?	No
Real journals named?	Yes
Real conferences named?	No

Appendix B – Interview Framework

eIUS/e-Uptake Interview Framework: Version 2

Duration: 1 Hour

Introduction (5 Min)

Give brief introduction to eIUS and e-Uptake, including overall aims.

Remind participant about informal/semi-structured nature of interview, and intention to tape interview, in addition to taking notes.

Remind participant about ultimate intention to publish findings as 'Experience Reports' – depending on participant's approval of the written up version.

Ask to sign consent form if haven't already and turn audio capture device on.

Interview - eIUS Questions (25 Min)

What percentage of your time do you spend on research?

Could you tell me briefly about your particular research area and give me a couple of examples of specific research questions that your research attempts to answer?

I'm interested in understanding the types of day-to-day research tasks that you carry out in order to help you move forward with this research agenda. I'm interested in doing this in the context of a generic type of research cycle from literature review, data collection/processing, all the way to publishing. So if I can start by asking you about … ?

Literature Review
- Media: Digital or Print
- Type: Journal, Book, Preprint, Report, Thesis

Collect Data
- Primary data, Secondary data, Experimental data, Simulation data

Process Data
- Statistical analysis, Qualitative analysis, Critical analysis, Annotation

Discuss (link findings to prior work)
- Face to face, Phone, Email, Discussion List, IM (1:1, Chat Room), Audio conference, Video conference, Shared documents

Publish
- Media: Digital or Print
- Type: Journal, Conference Paper, Report, Book, Thesis

During questions relating to these tasks, enquire as to which facilitating tools/services are used. Also ask, 'How essential is this tool for your research – would you be able to do the research without this tool?'

Interview - Service Type Check (5 Min)

Scan (privately) check-list of service genres - note (to self) any major areas that haven't been covered. Ask participant about these areas.

Interview - e-Uptake Questions (15 Min)

1. Where people mention barriers/problems: dig deeper by asking what was the root cause of the problem and (how) was it overcome? (no need to do this immediately, can go back to something that was said earlier) If not, ask at the end: what problems have you encountered in your work, what was the root cause in these and (how) were they overcome?

2. Looking at any problems encountered, how would you classify them, e.g., were they primarily technical, organisational issues, legal issues, etc.?

3. Thinking more widely, what do you consider to be the main barriers to wider uptake of e-infrastructures in your community?

Break (5 Min)

Offer refreshments.

De-Brief (5 Min)

1. Ask them how they felt the organisation of the interviews went – is there anything they feel could be improved.

2. Go through consent form - ask them if it made sense.

3. Go through the interview framework asking them how they thought it went.

4. Ask researcher about benefits they felt they would get from participating, and what is the number 1 benefit to them. Give examples:
- Opportunity to publicise successful outcomes of the use of services and potentially inspire others through, for example, co-authored publications with minimal effort on the part of the interviewee.
- Networking opportunities (career progression)
- Having an impact on UK service providers, i.e. improving the services.
- Reflecting on, and sharing research methodologies.

5. Ask them if they know anyone else particularly in other institutions who might be good interview candidates.

6. Ask for any supporting documentation.