



Steven Newhouse, University of Southampton
Jennifer M Schopf, National e-Science Centre and Argonne National Laboratory
Andrew Richards, Council for the Central Laboratory of the Research Councils
Malcolm Atkinson, e-Science Institute

Abstract:

SUPER, a Study of User Priorities for e-infrastructure for Research, was a six-month effort funded by the UK e-Science Core Programme (EPSRC & JISC) to identify short-term needs that can be delivered within existing funding streams from organisations, and longer-term concerns. The aim is to provide a usable, useful, and accessible e-infrastructure for researchers across a wide variety of disciplines, integrate existing and new e-infrastructure facilities and services into a coherent whole, and provide the basis to increase the use of the existing e-infrastructures by a factor greater than ten by 2010.

The study involved visits to seven universities, and interviews with approximately 45 people active in more than 30 projects. This report sets out the major recurring issues that were identified in these discussions, and proposes a detailed programme of investment in software, policy and support in order to address these needs and concerns.

UK e-Science Technical Report Series

Report UKeS-2007-01

Available from http://www.nesc.ac.uk/technical_papers/UKeS-2007-01.pdf

Copyright © 2007 The University of Edinburgh. All rights reserved

Executive Summary

SUPER, a Study of User Priorities for e-infrastructure for Research, was a six-month effort funded by the UK e-Science Core Programme and JISC to identify short-term (6–18 months) needs that can be delivered within existing funding streams from organisations, and longer-term (3–5 years) concerns. During the study we visited seven universities and spoke with approximately 45 people active in more than 30 projects.

From these discussions we identified five major recurring common issues:

- Distributed file management and policy relating to their curation
- Tools to support the creation and use of dynamic virtual organisations
- Need to support projects through tools and training
- Operational provision of authentication, software licensing, and reliable and consistent environments across distributed resources
- User interaction with e-infrastructure services through defined APIs, scripting environments, and graphical workflow environments

As a result of the survey we recommend investment in three broad areas: *software*, *policy* and *support*, with items listed in no particular order.

- Software
 - Automatic data annotation and provenance tools to support domain specific schema
 - Mechanisms to support controlled and convenient sharing of files between groups
 - Reliable documented software base to enable virtual organisations built around individuals to gain access to services and resources, and collaborative mechanisms to facilitate research between these individuals
 - License- management software, across a site and across a VO
 - Simplified authentication tools – better user interfaces and focus on ease of use and local (client) deployment
 - Integration of e-infrastructure resources into the environments in regular use by applied end-users (e.g., Matlab, Jython, Perl, R, portals)
 - Better tools for understanding failures and increasing reliability
- Policy
 - Development of a best practice document to support research groups in developing their own data curation and file management policy
 - Development of common annotation schemes for individual communities to enable consistent metadata labelling within these communities
 - Defining a set of recognised roles, actions, and relationships to support the construction of virtual organisations
 - Consistent environment guidelines for machine interactions, especially with the NGS, so users see one environment across a Grid
- Support:
 - Better technical consultancy to end-users who wish to move their applications to use e-infrastructure services, developers who wish to use best practice to build e-infrastructure services, and deployers who need to

configure e-infrastructure services for maximum performance and reliability – in general, better information with a human behind it (this needs to be a funded, coordinated service with experts behind it – simple lists are worse than nothing, as they are always dated and often misleading)

- Provision of self-help training materials and hands-on tutorials delivered to users throughout the UK through national and local trainers for common tools and software being used within the UK community
- Outreach and education about what is possible to non-traditional communities

This work also had two additional sources of data. On February 16, 2006, we ran a workshop that gave an overview of the results of a draft of the report. A workshop summary is given in Appendix B. In addition, we ran an online survey form for the broader community; results from this survey are summarised in Appendix C.

The UK e-Science Centre Directors' Forum was consulted by email and by discussion of a draft of the report at the Forum meeting held in the Reading e-Science Centre on the 19th January 2007. The report received general endorsement at that meeting.

Contemporaneously with the conduct of this study, the results of the Office of Science and Innovation working groups on e-Infrastructure were published (<http://www.nesc.ac.uk/documents/OSI/index.html>). These take a longer-term view of UK e-Infrastructure than the SUPER study and they did not link as closely with *current* users and services. Nevertheless, the findings of SUPER are consistent with early steps towards the long-term vision of those reports.

1 Introduction

SUPER, a Study of User Priorities for e-infrastructure for Research, was a six-month effort funded by the UK e-Science Core Programme (EPSRC & JISC) to inform investment so as to

- provide a usable, useful, and accessible e-infrastructure for researchers across a wide variety of disciplines,
- integrate existing and new e-infrastructure facilities and services into a coherent whole, and
- provide the basis to increase the use of the existing e-infrastructures by a factor greater than ten by 2010.

E-infrastructure encompasses the facilities and services that enable more advanced or more effective research through support of access to information, data, computation and distributed collaboration. Since multidisciplinary teams often span national as well as institutional boundaries, support for international collaboration and consistent provisioning must be considered. Inevitably, multiple organisations will be engaged in constructing and operating e-infrastructure; therefore the recommendations must permit autonomous providers to collaborate in order to provide coherent and consistent facilities.

As a first step in carrying out the SUPER project, we spent several months in late 2006 meeting with approximately 30 groups across the UK who represent a cross section of the research community and who were engaged in a variety of roles and projects. The groups interviewed included current and potential *end-users* of e-infrastructure, generally those conducting research, design, analysis or diagnosis in projects funded by one of the UK Research Councils; *technologists*, or developers who take generic middleware and adapt it for a specific use case; *generic tool developers*, those who were building solutions that could be used in many application domains; and *service providers*, including universities and publicly funded research institutes that are setting up hardware and services.

Our meetings lasted from half an hour to half a day and covered a wide variety of topics, concentrating on current use and needs. We considered performing structured interviews based on a standard questionnaire, but differences in backgrounds and knowledge of those interviewed made this approach ineffective, so the interviews were not formally structured as such. Instead, we asked what functionality the groups had tried in the past, what their applications needed today from the current Grid infrastructures, and what functionality the groups were considering for near-future plans. Most meetings ended by our enquiring what functionality the group thought was most important but lacking in today's tools or services. We were interested in both short-term (6–18 months) needs that can be delivered within existing funding streams and longer-term (3–5 years) concerns that may need to be resourced within new funding streams.

In terms of coverage, we performed face-to-face interviews over the course of the Autumn 2006 in Newcastle, Glasgow, Edinburgh, Oxford, Cambridge, UCL, and Reading, covering 45 people from over 30 projects. Figure 1 shows how these divided according to funding project, and a full list can be found in Appendix A.

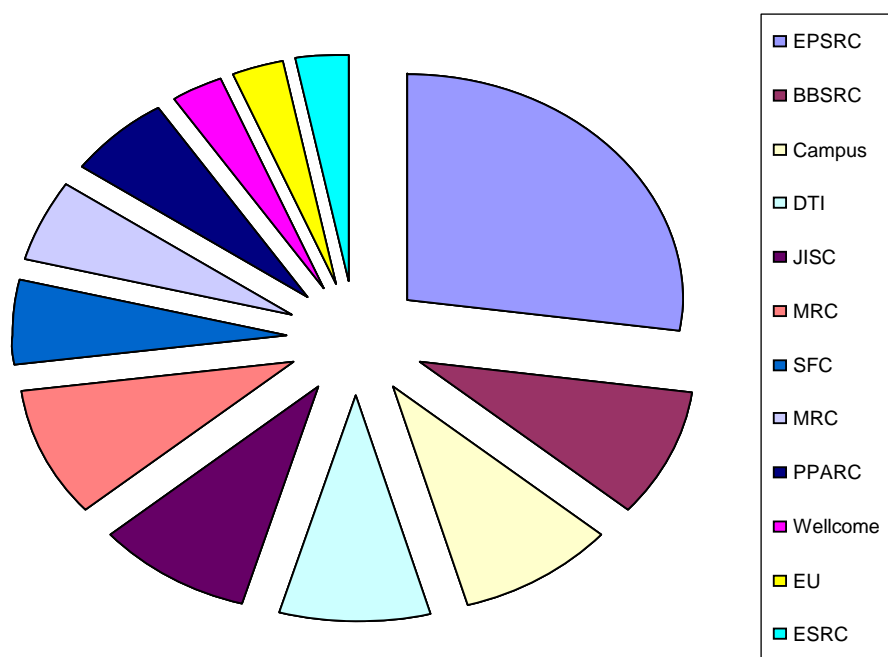


Figure 1: Funding for projects that were interviewed.

We then collated the responses and synthesised the significant common requirements across these groups. For the purposes of discussion we have categorised these requirements as “major,” discussed in Sections 2–6, and “minor,” discussed in Section 7. We have not included in the report other issues that came up in our discussions but were associated with only a particular community. In Section 8 we provide a set of recommendations and identify where they can be supported within existing funding streams (e.g. OMII-UK, NGS, DCC, JISC e-infrastructure) and where future funding needs to be sought to support these activity. In general, however, we have not identified “best” technical solutions. Such a task will require additional detailed requirements gathering and analysis that is beyond the scope of this broad study.

The interviewees were given a chance to comment on an early draft of the report, as were the UK e-Science Directors’ Forum.

This work also had two additional sources of data. On February 16, 2006, we ran a workshop that gave an overview of the results of a draft of the report. A workshop summary is given in Appendix B. In addition, we ran an online survey form for the broader community; results from this survey are summarised in Appendix C.

This report is intended to supersede previous reports [1, 2] as a guide to prioritise effort and investment. Our goal is not to suggest technologies or solutions to these issues. We expect solutions to emerge through community consensus following the publication of this report.

2 Collaborative File Management and Policy

The early vision of Grids – running applications on remote shared distributed resources ranging from dedicated clusters to shared-use, cycle-stealing desktop machines – is in use in many communities. The availability and accessibility of these resources, as a side-effect of the UK e-Science programme, have brought new user communities to larger-scale science than was previously feasible, and have even promoted paradigm-shifting science techniques in some domains such as large-scale ensemble and parametric studies.

By far the largest concern of the users we spoke with, centred on the use of large-scale data stored in structured file sets – *how to share data with colleagues* within their project or their wider community. Of particular importance is the management of data stored in files, whether it was software, results or other data, as opposed to data stored in databases. Users are concerned about the data’s long-term storage and curation, about means for accessing these files from the local desktop, and about seamless data transfer to remote resources for further analysis. End-users also want to be able to annotate the files with metadata about the contents and provenance, in order to enable searching and reanalysis at a later date – *metadata is the key to being able to share the results*. Many of the groups have explored solutions such as the Storage Resource Broker (SRB); however, SRB is seen as a heavyweight solution that was hard to deploy and maintain by a project.

Research has experienced a paradigm shift with respect to the changing uses of data as well as the changing use of compute resources. More groups now need to share data more widely under defined access policies, to merge data, and to retain that data longer. There was an identified need for *standard policies and tools for data curation* across RCUK-funded projects. Those policies are currently not well defined, but they need to be – for both user roles and temporal constraints. For example, several groups mentioned a “guarded period of access,” during which only a limited group of people could see the data and then wider access to the community after a set time, perhaps upon publication of results.

For some groups “wider access” implies being open to the community only, perhaps within a well-defined virtual organisation or set of virtual organisations, whereas for other groups it means being readable to anyone. Some groups have quite strict access limitations, including what resources were even able to house the data. Any access control structure and its underlying authentication mechanisms must support controlled access to any external collaborators and eventually unrestricted access to others. But in general, there is a shift towards much *longer-term storage of data*, some for pragmatic experimental use, and some at the behest of the funding agencies.

Easier *access of data* was also an issue. Many groups now have to manage the file output from computations across multiple locations, including national resources such as the National Grid Service (NGS), as well as campus and desktop resources. Researchers would like to seamlessly access their local files when running an application remotely, so that they can edit input files locally that form the basis of the simulation. Likewise, the output files residing on a remote resource need to be accessible for future processing and analysis on the local desktop or on other remote resources. This situation leads to requirements for the registration and discovery of files held in different locations.

In order to be able to easily curate and annotate data files, additional tools are needed to *autogenerate metadata* about the data itself and how, where and by what that data was generated – its *provenance*. Once provenance data is collected there will also be a requirement to navigate and analyse the data. It was noted that if users are responsible for the annotation of their data the task generally is left undone. A separate issue concerns how the data is annotated. Many groups have begun to create their own metadata after failing to find acceptable community standards. Standards exist for common basic properties, sometimes many of them, for example in the instance of timestamps, and users were generally happy with the existing frameworks for this higher-level information. However lower-level and domain-specific metadata has yet to be standardised for many communities. The quality of such annotations is often uneven, however, and some communities therefore have started developing common metadata schemas to drive annotation standards. Automated collection of basic metadata is seen as a large step forward from current practices; but for some domains, specialists still may be required to do the full job by hand; almost always, some human input is required, e.g. the purpose of an experiment.

3 Tools to Support Dynamic Virtual Organisations

Although technical aspects dominated many of our discussions, there was a growing recognition of the importance and difficulty in building an effective multidisciplinary multi-site virtual organisation (VO) within a project. This was also the topic of a workshop that we attended [3]. Established management practices need to be adapted to deal with the challenges of a virtual organisation, where the incentives for individuals may be radically different and not under the control of the project management.

Part of the lack of established practices for a VO is due to the fact that there is currently no clear definition of a VO. Undetermined factors include how dynamic the VO is and if it involves changes in membership of people or in resources. Some VO's are being defined top-down, with centralised enrolment, approval, and resource allocations that may have relationships reflecting off-line legal obligations and strong monitoring of service level agreements. Other VO's have a bottom-up approach, generally reflecting more social network groups with role-based agreements.

While current virtual organisations layered on top of production Grid infrastructures are able to adapt to a dynamic set of resources, it is harder for the virtual organisation itself to be dynamic and to bring in new people or resources that are not already part of a larger collaboration. *Support for dynamic virtual organisations* oriented around people and real organisations is becoming critical for future collaborations and controlled resource sharing (services, data, etc.) but varies considerably between domains, for example, business to business, accessing services in a production Grid, or more spontaneous collaborations based around social networking.

Clearly, the virtual organisation model needs to evolve to *more than just a management of roles*. Relationships, obligations, actions and interactions within a distributed team are complex and will change over time. These relationships will also have different levels of enforcement, varying from best effort to legally binding contracts. In addition, services and tools are needed not only to enrol, authorise and end an individual's participation in a virtual organisation, but also to manage the

relationships and to register transactions delivered as part of this relationship, in order to provide a non-repudiable record of the interactions.

The tools and services need to reflect the different stakeholders, such as the management of the virtual organisations, service providers, developers and end-users, so as to provide a “backplane” to support collaboration. Most VO tools are built for system administrators and the interfaces are suited to these technically experienced users. But the end-users may simply want to know what resources they have access to, without in-depth information about certificate setup procedures or accounting requirements. Currently, there is a *lack of tools oriented towards the end-user* in this space.

For the few tools that exist, usability remains a big issue. Current tools, in part because of a lack of consistency in what is meant by a VO in different settings, often solve problems that aren’t what the user is interested in and the tools are often very difficult to understand or use in production settings.

Moreover, the current support for *collaboration within virtual organisations* is extremely limited. Many of the tools available for managing virtual organisations in a production environment are considered hard to use, have a static structure with inflexible operations, or are driven by the needs of the service providers rather than by those trying to access these resources. A successful virtual organisation needs collaborative tools that extend beyond access control and accounting. Mechanisms such as instant messaging have become widespread in order to support a dispersed project team, but the ability to share other resources such as workflows, services and documents is missing. Audio and video conferencing tools are seen as essential or desirable by many groups, but there is general disappointment in the overheads and ineffectiveness of the currently available tools. Although the need for these tools is being driven by end-users, and although these tools can be installed by end-users, training is still needed for system administrators to ensure that sufficient bandwidth is made available and networks are configured appropriately.

Without a common understanding of what is meant by a VO and the entities within it and their relationships, it is very hard to develop tools that can be reused across different VO’s. Tools should be developed that utilise these defined relationships and entities to enable flexible and dynamic construction of VO’s. Once a common vocabulary has been established for VO’s, different VO’s will be able to share resources and interact in an organised manner. Such activity should not expect a single VO model; indeed, we encountered as many VO models as projects in our discussions. This is an area where standards, such as XACML, are beginning to gain widespread use. As work progresses in this area, it is important to keep in mind that tools need to be flexible for different communities, we need to respect diversity in how VOs work, and to be careful to not force VO’s into one mode of operation that may not be suitable for them.

4 Project Support: Teams, Services, and Training

A recurring theme in many of our discussions was the importance of an effectively organised team to a project’s success. To meet the demands of multiple stakeholders, a balanced, trained, multi-skilled team is critical to the development and deployment of complex software infrastructures.

The *skills and roles* needed by such a team, and therefore its membership, will likely vary over time. For instance, more software development effort might be needed during the early phases of a project, with more effort on deployment and training towards its end. In the early phases of the UK e-Science programme expertise in the emerging middleware infrastructures was brought into an applied science project through collaboration with computer scientists, e-Science centres or other specialists. Some of these skills are now being brought into a project through training, collaboration with external groups or national organisations, or new personnel. Nevertheless, several groups commented on how hard it is to find such personnel and bemoaned the lack of available specialists in the necessary fields (e.g., Web services, HPC programming, application performance tuning and Java programming).

Several projects have learned the hard way that it is easy to spend the majority of their time building new middleware and support services, and not getting their end-user science done. Hence, there is a much *broader acceptance of third-party software and services* than two years ago [2]. Even groups that had started out building their own infrastructure are turning to pre-packaged solutions for higher-level services when possible. However, there is a *notable lack of information* about what is truly available and what tools might be better suited to a particular use case. Several groups are making assumptions about the software available and its state; assumptions that may not be well informed. Even those groups that have basic information about what tools and services are available may not know how to make sound decisions when faced with a choice. To provide such expertise is, of course, an overwhelming problem – no one practitioner can advise across the many areas of Grid software and services. One promising solution is to use “vendors” such as the Matching Agency [4], offered by the Arts and Humanities e-Science Support Centre, which matches applications with software experts for advice and future work.

Experiences from the surveyed projects also show that no single software provider provides the *end-to-end solution* needed for most successful e-infrastructure deployments. While this is not necessarily a problem from an end-user’s perspective, it means that projects deploying their own infrastructure need the human resources and skills to integrate these components from different providers. In order to avoid these overheads, enabling services are needed that provide access to key resources (e.g., compute, data, file movement). Infrastructure providers, such as the NGS and local campus Grids, play a key role in providing these low-level services. However, these are unlikely to be of direct use to particular domains or projects. In order to promote uptake into applied communities and projects, additional generic and domain specific services and tools are needed to bridge the gap between the e-infrastructure providers and end-users, such as the Grid-enabled access to data sets provided by MIMAS and EDINA. These higher-level, domain-focused services are easier to integrate into projects and, when hosted by “professional” service providers, provide a solid base for communities to build on. These services enable the software outputs from research projects to be consolidated. As communities build around these services, we would expect them to form the basis of standardisation within relevant forums, and subsequently part of the expected e-infrastructure service environment, as captured within roadmaps such as ESFRI [5] and JISC (www.jisc.ac.uk/aboutus/strategy/strategy0709.aspx).

Information relating to the availability of tools and services was not being effectively disseminated to many of the groups we talked to. When groups were aware of the various technology options they felt there was no expertise they could call upon to help them in their decisions. Communication between generic providers and end-users continues to be hampered by the fact that the *groups communicate quite differently*. This situation is also complicated by the fact that different members of the end-user group will want different end tools, a fact that is not always clear to the tool developer. A better understanding of how to manage dynamics between these groups is needed.

At a higher level, today's e-infrastructure users go well beyond the risk takers who started work in this area and as the end-user community expands, there arises a need for *broad outreach and evangelising* of the benefits provided by adopting the emerging campus, national and international e-infrastructure deployments. Many new projects simply don't know what is possible or probable; and many new communities don't know what is easy and available versus hard or even an open research question. There is no baseline of capabilities agreed upon in the broader community, and this situation is significantly hampering additional uptake.

In addition to promoting the adoption of available software and services, *training* needs to be provided for the different stakeholders within a project – the end-users who will be using the e-infrastructure services through the tools and applications, the developers (both of generic and domain specific services) who will be using the deployed e-infrastructure services to build other services and tools, and the deployers (within both a local and national context) who will have the responsibility of managing the required e-infrastructure. Training materials are needed in many forms: formal instructor-led courses, standalone self-help material, worked examples, reference systems, and so forth. Ironically, many infrastructure projects noted the need for training and had funds – but were ignorant of existing training materials (from software providers) or ongoing courses (from organisations such as NGS or NeSC) – and as a consequence there was considerable duplication of activity.

Although training can accelerate skill transfer into a team, it cannot provide experience. Several teams expressed a desire to be able to tap into external expert advice as required, rather than build up such experience internally. Advice on how to best use the deployed software needs to continue beyond high-quality training and self-help courses with access to experts who can provide consultancy on best practices and application related problem solving. In addition, support for working with legacy codes is needed. Advising, in general, may include embedding specialists with applied application groups to ensure an effective two-way dialogue that not only disseminates expertise but also captures detailed requirements to guide future developments (although, admittedly, the scalability of this approach must be considered).

5 Operational Issues: Authentication, License Management, and Reliability

Many user communities, while technically capable in their own fields, are not interested in dealing with the complexity of Grid middleware, especially their authentication systems.

Authentication remains a concern for most of the groups we spoke with. The adoption by most service providers of X.509 public key certificates as the primary authentication mechanism for accessing services raises many concerns about end-user usability, but not performance. There still exists a strong need for *simple authentication mechanisms at the point of use*, which may include hiding all certificates from end users. Policies for certificate issuing are also becoming a problem as collaborations extend beyond national boundaries. In general, none of the groups we spoke with using certificates are satisfied with their current solution.

Issues with licenses were much more common than two years ago as a result of the broadening of the Grid user base into more applied areas. In many current deployments, strongly licensed software is used because it is the community standard and because good, open source alternatives are not available and are unlikely to emerge, given the dependence of these communities on verified, quality-assured applications. However, the number of licenses available for a given piece of software is limited, and *currently there isn't a good solution for managing the shifting of licenses* on machines across a given site, let alone within a full VO. For those projects with industrial collaborations, this problem is even more significant because if a nonacademic partner uses academically licensed software, both groups can get blacklisted, and then many thousands of pounds will need to be spent to access the software.

Given the ability to identify the user and to provide access to a suitable license, the next step is to successfully provision the application on the remote resource. Here, the *need for a consistent environment* was commented on by groups using different resources. Re-establishing one's personal user environment on each of the NGS sites is a simple example of a current impediment to use. Virtualisation technologies are one approach to providing consistent portable environments; however, these are not currently in production use. Moreover, none of the tools in current use have client toolkits that are considered simple or straightforward enough for today's users. Indeed, a common complaint was that, even when the tools were installed, there were ongoing problems with getting the environments to be what was expected. There also was agreement that we need to move away from the remote login approach, but the tools available are simply not stable and reliable enough for hard use.

In general, *reliability was still seen to be lacking* by most users we spoke with. Even with professionally hosted Web services, there seems to be a lack of stability with either constantly changing interfaces or endpoints. As a result, many groups simply do not expect that the services will be up; they just work with whatever they find responsive at run-time. The statement "Always *something* doesn't work" was repeated by several groups. At times, the software is at fault, in that remote errors are trapped and sometimes effectively dealt with by the remote service, but the error or recovery status is not propagated back to the client application invoking the software. At other times the service providers had made changes. Error messages are still found to be lacking, when they are delivered at all. The complicated stack of services also doesn't help when application or system debugging is needed. Even the monitoring software may give conflicting results.

6 User Interaction

Different communities (and users) use different environments to undertake their work. When offered the use of e-infrastructure services to accelerate their research, users frequently express reluctance to change environments solely for the purposes of accessing these new capabilities. It is therefore critical for increased adoption to understand how the different users (end users, developers, etc.) interact with e-infrastructure services.

In order to enable tool builders to integrate access to e-infrastructure services into existing user environments, there is a *need for stable APIs* to build on. Ideally, the APIs will be based on standards and not local to a single toolkit or implementation. Ultimately, these APIs need to encompass the full breadth of e-infrastructure services that are available, but initially they should focus on job submission and management, service discovery, file movement and management, and data annotation.

These standard APIs can be used as a basis for accessing e-infrastructure services from within nongraphical environments; alternatively, command-line clients can be executed from within existing shell environments. Over the years, many projects have invested considerable effort in developing a suite of scripts to support their interaction with Grid services by using scripting environments (such as Matlab, Python, Jython and Perl) and UNIX shells (such as Bash and Tcsh). Porting these scripts to a new middleware infrastructure, or even to a new release of the current software, represents a considerable risk to a project.

Many groups are transitioning to a workflow system, either to better track their intermediate results (see Section 2 for a discussion of data needs) or to keep track of a large number – often tens of thousands – of individual jobs. Most groups are considering an already running system such as Taverna or one of the BPEL implementations, as opposed to building their own from scratch or further extending existing scripts that had been the start of a home grown system.

Portals provide another approach for end-users who require a simple intuitive interaction model. *Portals have gained considerable acceptance in some communities.* Some projects have successfully exposed common operations involving e-infrastructure services to a predominantly non-technical (in a computing sense) community, while other projects have found the inflexibility frequently offered by a portal solution a hindrance. The availability of customisable portal toolkits (through the use of portlets) provides an important approach to building solutions for communities where they are needed.

7 Other Issues

In addition to the major issues discussed in Sections 4–6, several other issues arose, which we discuss here.

7.1 Sustainability

The move to the Full Economic Cost (fEC) funding models and the potential loss of investment from the Strategic Research Investment Fund (SRIF) is causing concern among many providers of research computing – especially at the campus level. A key requirement within such a costing environment is the integration of accounting tools

into the deployed services so that activity can be charged back to individuals and projects if required by local policy. These providers recognise that investment needs to be made in training and supporting their local users in moving to the new e-infrastructure, and they do not see this resource appearing at present.

7.2 File Replication

The issue of *replication* was mentioned by several groups in different contexts. Infrastructure providers stated a need for a national-level backup or mirroring service. Several groups with large data are looking at the QCD Grid approach to replication for their own large sets, especially to reduce access times and to improve availability. Replication was less of a concern for groups with smaller data, who felt in part that the policy overhead outweighed the advantages. The politics of replicating data was also noted as a hindrance to some.

7.3 End-to-End Security

End-to-end security (of the underlying resources and the data accessed through them) is clearly highly desirable when activities span different organisations. This is a nontrivial goal, however, and the approach taken by many service providers is deployment of *firewalls*, which continue to provide problems to end-users. For example, changes in firewall deployment cause unexplained errors for many Grid services, decreasing the reliability and use of the service. In addition, many sites are now strongly restricting the number of open ports allowed on a firewall, thereby making deployment of Grid software difficult even when the needed ports are well documented.

7.4 Scheduling

In general, users said they would like a single point of entry to resources with common environments if it had the same features that they were used to with something like PBS. Only one group we spoke with had a pressing need for advance reservation and co-scheduling, and that was currently being well addressed with the *ad hoc* systems in place. Having more predictable resources is seen as an advantage that a common scheduler front-end might be able to supply, but in general this seems to be a lower priority than other pressing data or security needs.

8 Conclusions and Recommendations

As a result of the survey we recommend investment in three broad areas: *software*, *policy* and *support*, with items listed in no particular order. Sustained investment in these areas will provide a set of structured tools, services and environments to support access to e-infrastructure, and a support infrastructure to enable the adoption of e-infrastructures by new user groups.

- Software
 - Automatic data annotation and provenance tools to support domain specific schema
 - Mechanisms to support controlled and convenient sharing of files between groups
 - Reliable documented software base to enable virtual organisations built around individuals to gain access to services and resources, and collaborative mechanisms to facilitate research between these individuals
 - License-management software, across a site and across a VO

- Simplified authentication tools – better user interfaces and focus on ease of use and local (client) deployment
- Integration of e-infrastructure resources into the environments in regular use by applied end-users (e.g., Matlab, Jython, Perl, R, portals)
- Better tools for understanding failures and increasing reliability
- Policy
 - Development of a best practice document to support research groups in developing their own data curation and file management policy
 - Development of common annotation schemes for individual communities to enable consistent metadata labelling within these communities
 - Defining a set of recognised roles, actions, and relationships to support the construction of virtual organisations
 - Consistent environment guidelines for machine interactions, especially with the NGS, so users see one environment across a Grid
- Support:
 - Better technical consultancy to end-users who wish to move their applications to use e-infrastructure services, developers who wish to use best practice to build e-infrastructure services, and deployers who need to configure e-infrastructure services for maximum performance and reliability – in general, better information with a human behind it (this needs to be a funded, coordinated service with experts behind it – simple lists are worse than nothing, as they are always dated and often misleading)
 - Provision of self-help training materials and hands-on tutorials delivered to users throughout the UK through national and local trainers for common tools and software being used within the UK community
 - Outreach and education about what is possible to non-traditional communities

We hope that these recommendations will influence the individual roadmaps and activities of organisations charged with supporting collaborative multidisciplinary science in the UK (e.g. OMII-UK, NGS, DCC) and their funding bodies – the UK research councils and JISC.

Acknowledgements

We gratefully acknowledge all of the people we talked to; their time and frank discussions are appreciated. In addition we received valuable feedback and comments from Dave Berry (NeSC), Richard Sinnott (University of Glasgow), and David Wallom (University of Oxford).

References

- [1] Geoffrey Fox and David Walker, “e-Science Gap Analysis,” UK Technical Report Series # UKeS-2003-01, June 2003.
- [2] Jennifer M. Schopf and Steven J. Newhouse, “Real Users - Real Requirements,” UK eScience All Hands Meeting 2005, Sept. 2005.
- [3] Virtual Organisations and Grids: Applications, Patterns and Infrastructures Workshop, 21-22 Nov., 2006, e-Science Institute, Edinburgh [http://wiki.esi.ac.uk/Virtual Organisations And Grids](http://wiki.esi.ac.uk/Virtual_Organisations_And_Grids)
- [4] Arts and Humanities E-Science Support Centre Matching Agency, <http://www.ahessc.ac.uk/matching-agency>
- [5] ESFRI: European Strategy Forum on Research Infrastructures, <http://cordis.europa.eu/esfri/>

Appendix A: Projects Visited

CARMEN (EPSRC)

http://bioinf.ncl.ac.uk/carmen/index.php/Main_Page

Paul Watson (Paul.Watson@newcastle.ac.uk)

Phil Lord (Phillip.Lord@ncl.ac.uk)

Functional MRI analysis involving both storage (50 TB) and compute (for analysis) Grid. Bespoke analysis scripts built around C and Matlab libraries. Linking into SRB and NGS.

Gold (EPSRC)

<http://www.goldproject.ac.uk/>

Rob Smith (Rob.Smith@ncl.ac.uk)

Roles and responsibilities within virtual organisations.

CRISP (DTI)

<http://www.neresc.ac.uk/projects/index.php>

Rob Smith (Rob.Smith@ncl.ac.uk)

This project, led by Belfast e-Science Centre, is using Gold as a toolkit and other technology to support the scenarios within the CRISP consortium in order to drive the further development of Gold. Focussing on expressing the relationship between the VO and other organisations.

ComparaGrid (BBSRC)

<http://www.comparagrid.org/>

Neil Wipat (anil.wipat@ncl.ac.uk)

Dan Swan (d.c.swan@ncl.ac.uk)

Matthew Pocock (matthew.pocock@ncl.ac.uk)

Interest is in workflows to do genome comparison using Taverna rather than service provision. But current services are fragile, so they deploy their own services (e.g., BLAST) to bring the reliability under their own control. Processing of workflows using large data files is an issue over SOAP (Axis1); the group members are moving to Cross fire & Axis2

OGC (JISC)

<http://www.opengeospatial.org/>

Gobe Hobona (G.E.Hobona@newcastle.ac.uk)

Providing a portal to a 'semantically aware workflow engine for geospatial workflows'. Data source is a Web service that provides access to a geospatial data set – in this project an Ordnance Survey dataset – but it can result in large query returns. JISC-funded Grid and Open Geospatial Consortium Collision project with EDINA.

NanoCMOS (EPSRC)

<http://www.nanocmos.ac.uk/>

Asen Asenov (A.Asenov@elec.gla.ac.uk)

Campbell Millar (C.Millar@elec.gla.ac.uk)

Scott Roy (S.Roy@elec.gla.ac.uk)

Richard Sinott (ros@dcs.gla.ac.uk)

Antony Stell (ajstell@dcs.gla.ac.uk)

Dave Berry (daveb@nesc.ac.uk)

Project for “meeting the design challenges of NanoCMOS,” which includes device and circuit simulation using application expertise in different locations. The overall goal is to reduce the time to market through simulations. Identified needs involving security, steered workflows, metadata services, data management, licensing, and resource management.

GridQTL (BBSRC)

<http://www.gridqtl.org.uk/>

John Allen (jallen@nesc.ac.uk)

Jos Koetsier (jos@nesc.ac.uk)

Project dealing with mathematical algorithms to analyse genetics. The application is presented to the users through a portal run using GridSphere, which provides access to computational resources and data.

DGEMap (EU)

<http://www.dgemap.org/>

Richard Baldock (Richard.Baldock@hgu.mrc.ac.uk)

Yin Chen (yin@nesc.ac.uk)

Provides a European wide resource to access data on the development of the human embryo capturing through 3D imaging the genetic activity over time.

EDIKT2 – Chemistry (SFC)

<http://www.edikt.org/edikt2/>

Paul Madden (Paul.Madden@ed.ac.uk)

Andrew Turner (andrew.turner@ed.ac.uk)

A collaboration between universities at St Andrews and Edinburgh to provide an HPC infrastructure to enable the execution of ‘legacy’ computational chemistry codes with a requirement to provide file replication to support off-site backups.

EDIKT2 – Biology (SFC & Wellcome Trust)

<http://www.edikt.org/edikt2/>

Ilan Davis (ilan.davis@ed.ac.uk)

Alistair Kerr (alistair.kerr@ed.ac.uk)

Russell Hamilton (Russell.Hamilton@ed.ac.uk)

A project focused on managing the data generated from microscopes generating up to 500GB a year with potentially up to 10 GB per an image.

Edinburgh Centre for Bioinformatics. Medical Research Institute (MRC)

<http://www.bioinf.mvm.ed.ac.uk/>

<http://www.bioinformatics.ed.ac.uk/about.shtml>

Donald Dunbar (donald.dunbar@ed.ac.uk)

Service provision for a variety of research projects – ranging from Web-accessible databases to experiment planning and database design.

Integrative Biology (EPSRC)

<http://www.integrativebiology.ac.uk/>

Blanca Rodriguez (blanca@comlab.ox.ac.uk)

Development of an application to simulate the electrical behaviour of the heart. Part of the project's wider goal is linking multiscale and multiphysics models to produce a complete simulation of the heart.

NeuroGrid (MRC), GIME – Generic Infrastructure for Medical Imaging (DTI), CancerGrid (MRC)

<http://www.neurogrid.ac.uk/>

<http://www.cancergrid.org>

Sharon Lloyd (sharon.lloyd@comlab.ox.ac.uk)

NeuroGrid: Clinical trial support

GIMI: Ethics of data governance

CancerGrid: Clinical trial workflow

Because of legal/ethical restrictions, in many cases any functions analysing the data need to be shipped and executed where the data resides.

Integrative Biology Virtual Research Environment (JISC)

<http://www.vre.ox.ac.uk/ibvre/>

Matthew Mascord (matthew.mascord@oucs.ox.ac.uk)

Improve the experiences of scientists within the IB project through better use of technology (e.g., digital paper, collaborative visualisation).

Earthquake Engineering (EPSRC)

<http://www-civil.eng.ox.ac.uk/research/structdyn/presentations/uknees.html>

Tony Blakeborough (Tony.Blakeborough@eng.ox.ac.uk)

Support to deploy a UK NIEES node to simulate the testing of a model bridge.

Oxford Internet Institute (ESRC)

<http://www.oii.ox.ac.uk/>

Bill Dutton (william.dutton@oii.ox.ac.uk)

Marina Jirotko (Marina.Jirotko@comlab.ox.ac.uk)

Issues in team building and organisational development and motivation; capturing technical and nontechnical requirements from the stakeholders within a project.

Ownership of IP and confidentiality issues vary between different research councils.

Oxford Campus Grid

<http://www.oerc.ox.ac.uk/>

David Wallom (david.wallom@oerc.ox.ac.uk)

Amanda Barnard (amanda.barnard@materials.ox.ac.uk)

Example use of Oxford Campus Grid for simulation of nano structures that require large memory parallel computational resources. The staff within the OeRC provide user support and porting expertise to new users and those wishing to change what they are already doing on the resources available including university and NGS resources.

Reading e-Science Centre (EPSRC)

<http://www.resc.rdg.ac.uk/>

Jon Blower (jdb@mail.nerc-essc.ac.uk)

Multidisciplinary research centre specialising in the environment. Runs a DTI-funded project in collaboration with the Met. Office & IBM Delivering Environmental Web Services (DEWS).

Centre for Computing in the Humanities (AHRC)

<http://www.kcl.ac.uk/schools/humanities/cch>

<http://www.ahessc.ac.uk/>

Stuart Dunn (stuart.dunn@kcl.ac.uk)

Support for the A & H community in the adoption of e-infrastructure services.

NERC Data Grid (NERC)

<http://ndg.badc.rl.ac.uk/>

Bryan Lawrence (B.N.Lawrence@rl.ac.uk)

Provision of an infrastructure to support science; not a research project in its own right.

UCL Research Computing

Clare Gryce (c.gryce@ucl.ac.uk)

Jeremy Yates (jyates@star.ucl.ac.uk)

Central provision of research computing services - HPC, Windows-based Condor pool and storage. User and applications consultancy and outreach programme.

Infrastructure funded through SRIF 1 to 3.

Crystal Structure Prediction (EPSRC)

<http://www.cposs.org.uk/>

Sally Price (s.l.price@ucl.ac.uk)

Computational infrastructure to support research into organic crystal structures. Using automated workflows to produce a database of structures by executing analysis over a Condor pool.

OMII-BPEL (EPSRC)

<http://sse.cs.ucl.ac.uk/omii-bpel/>

Wolfgang Emmerich (w.emmerich@cs.ucl.ac.uk)

Development of BPEL environment funded by OMII-UK to support workflow in applied projects.

GridPP (PPARC)

<http://www.gridpp.ac.uk>

Ben Waugh (waugh@hep.ucl.ac.uk)

Support for data analysis from high energy physics at CERN.

AstroGrid (PPARC)

<http://www2.astrogrid.org/>

Jeremy Yates (jyates@star.ucl.ac.uk)

Provision of a virtual observatory using open source software for UK and international astronomers.

RealityGrid (EPSRC)

<http://www.realitygrid.org/>

Peter Coveney (P.V.Coveney@ucl.ac.uk)

Interested in running large-scale simulations on today's production Grid resources (e.g., NGS, TeraGrid, DEISA) – generally using Application Hosting Environment to abstract across different Grid middleware.

Cambridge Campus Grid (University of Cambridge)

<http://www.escience.cam.ac.uk/>

<http://www.escience.cam.ac.uk/projects/camgrid/>

Mark Calleja (M.Calleja@damtp.cam.ac.uk)

Mark Hayes (mah1002@cam.ac.uk)

A Grid infrastructure (primarily Condor based) that currently uses shared use departmental resources and in the future will also utilize dedicated clusters. The staff provide a support and consultancy service to application groups around the campus.

GROWL (JISC)

<http://www.growl.org.uk/>

Peter Brorsson (pb396@cam.ac.uk)

Provides access to a Web service (which acts as a proxy to other Grid middleware) from environments such as 'R'.

CCPNGrid (BBSRC)

Wim Vranken (wim@ebi.ac.uk)

Automatic calculation of protein structures from NMR data via a portal available to research scientists.

AstroGrid (PPARC)

<http://www2.astrogrid.org/>

Nic Walton (naw@ast.cam.ac.uk)

Services to provide a virtual observatory for UK scientists that can access resources in other countries through community standards within the IVOA (International Virtual Observatory Alliance).

Appendix B: Workshop Summary

In support of our interviews and this report, a workshop was held at the UK National e-Science Centre on February 16, 2007, to discuss the draft report and offer clarifications, which have been included in the main body of the report. The participants at this workshop were slanted toward policy makers and project management, e-Science strategists and e-Infrastructure providers. This contrasts with the focus on direct users that was chosen for the interviews.

The workshop itself consisted of an introductory discussion of the SUPER interviews and report, followed by breakout sessions to discuss in more detail the data, VO management, and support. Each breakout session was asked to consider the following points:

- Who are the early adopter/active communities?
- How uniform are the requirements within the community? Are there gaps? Do we need to revise emphasis?
- What are the targets over the next 12 months and then for longer term?

Talks and summary slides are available online at

<http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=743>

B.1 Data Breakout

The data breakout specifically focuses on the mechanisms, policy and replication issues with file & metadata management. In addition to the requirements in the main report, this group suggested the following actions:

- An evaluation of commonly available network file systems (GPFS, PVFS, etc) in comparison to distributed file management tools (SRB, SRM, dCache, etc)
 - First step would need to be an assessment of the criteria, for example, collaboration, deployment, ease of use, cost, etc
 - It would be important to communicate with additional communities, including HPC and DCC
 - This could be followed on by a workshop on available tools
 - This was proposed as a short-term need, that ETF or NGS might look into
- Best practice document for current metadata and annotation practices, and possible policies
 - In order to contribute data you have to have a policy for how annotation and metadata will be performed, preferably with standards to allow for interoperability
 - It was suggested this was a medium-term goal, with the most likely course of action being a JISC call for work
- Documentation of what standards are available for successful data curation
 - This could be followed on by a DCC workshop on metadata standards
 - It was noted that merely a list would be insufficient – lists are always dated before they're completed
 - This was felt to be a medium-term opportunity, likely to be addressed by the DCC with JISC
- A workshop or further outreach to better understand why some common tools available in the digital repositories space (e.g. from the DCC) are not in common use by e-Scientists.

- This was felt to be a short-term opportunity, that the DCC might be best to address
- Links page for already completed evaluations of institutional repository systems
 - Many institutional repositories have been evaluated, but this information is not easily available
 - A common listing that could be updated overtime, and a list of comparable criteria to use on future evaluations would be helpful
 - This was felt to be a medium-term objective, that JISC would be best responsible for
- Survey to understand why open source or commercial solutions for distributed file systems aren't in more common use
 - This short-term objective was felt to be the purview of ETF or JISC
- Survey To understand if many users are using data bases and we simply don't know it since they didn't mention it
 - Discussions during the SUPER interviews were very strongly file focused
 - This medium-term objective is likely to be undertaken by NeSC

B.2 VO Breakout

The VO breakout focused their discussions on the need to work on lowering barriers to entry, especially in terms of authentication. There is a perception that e-Science is heavy weight, e.g. Access Grid, so we should think about interactions especially first contact and how tools and projects should be marketed. Users should be made aware that e-Science isn't just about data and compute, and it doesn't mean you have to be doing something "big", but that it is a way to allow you to do your research and share your results more broadly and more easily.

This group identified a set of immediate and medium-term targets in addition to what was previously identified in the report. Immediate work included:

- Need to work on lowering authentication barriers to entry, e.g. Remove certificates and provide integrated single sign
 - JISC – Ann Borda to follow up.
- Marketing to change perceptions, e.g. e-Science is heavyweight
 - Malcolm Atkinson, Alex Voss to follow up.
- e-Research adoption theme to address at its workshops?
 - Alex Voss to address

Medium-term work included:

- User Enrolment and management tools
- Toolkits to enable integration
 - With environments and tools
- Social shaping to look at issues around VOs
- Development of concrete use cases to inform planning of future work
- Results of JISC e-Infrastructure work on VOs

B.3 Support Breakout

The support breakout group quickly identified that their clients, so to speak, were general e-Research communities. The support goal was currently to assist with training for specific technologies, but also to advise, for example, in the case of the

specific technology that a given PI should refer to in a grant proposal. We need to make sure the new generation of researchers is educated to use e-Infrastructure, and that we are enabling young scientists to have the skills they need for e-Research.

Short-term goals identified by this group, in addition to those mentioned in the report, included:

- Organise a federation of e-Infrastructure organisations within 12 months
 - The federation provides a friendly interface to our user community and connects each (potential) user to the appropriate members
 - Malcolm Atkinson, Ann Borda, Anna Kenway to follow up
- Better outreach machine: PhD/postdoctoral training centres, discipline specific centres, etc.
 - Malcolm Atkinson, Ann Borda to follow up
- Training events to make leaders in academia aware of the potential of e-Infrastructure for their research
 - Malcolm Atkinson, David Ferguson to follow up
- Discussions of how to step up as a national resource
 - Should we lead a push to get the funding councils to move away from their 'buy your own kit' model of funding?
 - More efficient cost models could be a persuasive argument
 - Malcolm Atkinson, Dave Berry to follow up

Appendix C: Online Survey Results

We posted a Zoomerang survey online from December 2006 to March 2007. It was advertised via over a dozen UK e-Science mailing lists and news letters in order to reach members of the community that we could not interview in person. We received between 24 and 26 responses for a large set of questions.

Table 1 shows the sources of funding of the responders. The funding sources mentioned as other were either university or EU-related funding sources. The project goals varied from infrastructure providers (hardware, data sources, digital curation, etc.), science and research groups (gene expression, engineering, networking, bio-informatics, high energy physics, etc.), to middleware, portal or visualization development groups. Table 2 shows the respondent's area of research, where each was asked to select all that might apply. Tables 3 and 4 show where the respondents felt their role was in a user taxonomy, as shown in Figure 2.

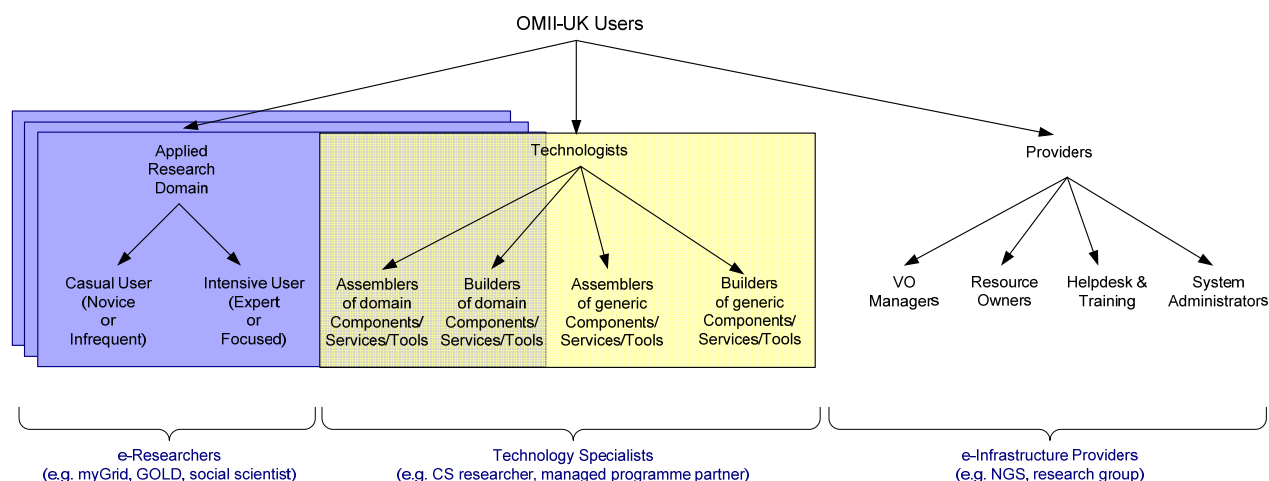


Figure 2: Taxonomy of users.

Tables 5 through 11 show the responses to a series of questions about the importance to the respondent for different facets of system software tools and services. Jobs are presented as being more important in this set of responses, but this doesn't contradict the interviews in that this question isn't where are you having problems, but how important is it to your project. As seen in the interviews as well, there is still a focus on the basic functionality (job submission before monitoring, file access before provenance or replication). And as expected, very few aspects of this problem were considered of no importance at all.

As a final question, we asked "What are the main issues, in your view, restricting your uptake of the e-Infrastructure?"

Responses could be roughly grouped as

- 1) With approximately 10 responses

- a. Complexity – many found the current infrastructure overly complex or unreliable
- 2) With 2-4 responses
 - a. Need for more basic instruction
 - b. Need for stable solutions, not upgrades.
 - c. Knowing that the resources and tools would be continued to be supported over time.
 - d. Simpler interfaces (e.g. portals) for basic functionality – matching tools to basic needs
 - e. Need for integrated admin support and common environments and interfaces across resources
 - f. Installation and maintenance is hard, and often not well documented
- 3) With a single response
 - a. Need for interoperability, especially in international projects
 - b. Need for more resources – UK available compute and data access not meeting project requirements
 - c. Hype has exceeded what is feasible today
 - d. Lack of understanding about the direct benefit to the end user communities – they want to do more science, write better papers, get more funding.

Table 1: Primary funding source by research council or funding agency.

AHRC	0	0%
BBSRC	2	9%
CCLRC	1	4%
EPSRC	9	39%
ESRC	1	4%
MRC	1	4%
NERC	1	4%
PPARC	4	17%
JISC	3	13%
Other, please specify	5	22%

Table 2: Respondent's research domain (multiple selections could be chosen).

Arts & Humanities	2	8%
Astronomy	5	20%
Atmospheric Science	1	4%
Biochemistry	1	4%
Bioinformatics	8	32%
Biology	6	24%
Cell Biology	1	4%
Chemistry	5	20%
Climate Modelling	2	8%
Computer Science	10	40%
Earth Observation	1	4%
Earth Science	2	8%
Economics	1	4%
Engineering	7	28%
Genetics	2	8%

Geography	3	12%
Grid Technology	9	36%
Humanities	2	8%
Integrative Biology	3	12%
Mathematics	1	4%
Medical	4	16%
Meteorology	1	4%
Oceanography	1	4%
Particle Physics	3	12%
Physics	3	12%
Plant and Microbial Science	1	4%
Social Science	3	12%
Statistical Analysis	3	12%
Telecommunications	0	0%
Other, please specify	2	8%

Table 3: Respondent's place in user taxonomy.

Novice/Infrequent User within a domain	2	8%
Expert/Frequent User within a domain	4	15%
Assembler of domain specific services/tools	3	12%
Builder of domain specific services/tools	4	15%
Assembler of generic services/tools	2	8%
Builder of generic services/tools	2	8%
VO Manager	0	0%
Resource Owner	1	4%
Helpdesk & Training	1	4%
System Administrator	1	4%
Other, please specify	6	23%

Table 4: Respondents role within user taxonomy.

Investigator	6	24%
Manager	7	28%
Worker	7	28%
Student	0	0%
Other, please specify	5	20%

Table 5: Importance of jobs/compute-centric services.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Job Submission	18 72%	6 24%	0 0%	1 4%
Job Monitoring & Tracking	7 29%	12 50%	4 17%	1 4%
Job Prioritisation	6 26%	7 30%	7 30%	3 13%

Table 6: Importance of data-centric services and tools.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
File Transfer Services	13 57%	7 30%	3 13%	0 0%
Data Access Services	16 67%	7 29%	1 4%	0 0%
Curation	7 30%	7 30%	5 22%	4 17%
Replication	2 9%	10 43%	7 30%	4 17%
Provenance	6 26%	12 52%	1 4%	4 17%

Table 7: Important of resource-oriented tools.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Resource Allocation	10 42%	11 46%	3 12%	0 0%
Resource Scheduling	8 33%	10 42%	6 25%	0 0%

Table 8: Importance of user management tools and services.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Account Management	7 29%	11 46%	6 25%	0 0%
Authentication (using Digital Signatures)	8 35%	8 35%	5 22%	2 9%
Authentication (using Username & password)	10 42%	8 33%	5 21%	1 4%
Authorisation (previously Security Infrastructure)	9 39%	7 30%	7 30%	0 0%
Auditing	2 8%	11 46%	9 38%	2 8%

Table 9: Importance of support services.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Training & Tutorials	11 46%	10 42%	3 12%	0 0%
Documentation	13 57%	10 43%	0 0%	0 0%
Technical Support	11 46%	10 42%	3 12%	0 0%

Example Applications	12	8	4	0
	50%	33%	17%	0%

Table 10: Importance of infrastructure services.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Monitoring Services	9 38%	10 42%	5 21%	0 0%
Failure Logging Services	7 29%	11 46%	5 21%	1 4%

Table 11: Importance of tooling.

Top number is the count of respondents selecting the option. Bottom % is percent of the total respondents selecting the option.	Very important	Important	Of slight importance	Not important
Failure Identification Tools	4 17%	14 58%	4 17%	2 8%
API Toolkits	7 29%	12 50%	2 8%	3 12%