ESRC e-Infrastructure: Portal Requirements

Rob Allan and Xiaobo Yang
STFC e-Science Centre, Daresbury Laboratory,
Daresbury, Warrington WA4 4AD

Rob Crouchley
Centre for e-Science, C Floor Bowland Annexe, Lancaster University, Lancaster LA1 4YT

Contact e-Mail: r.j.allan@dl.ac.uk, x.yang@dl.ac.uk, r.crouchley@lancs.ac.uk

Abstract

This report brings together conclusions from a variety of activities which have indicated requirements from the e-Social Science community (among others actively engaged in research) which can be satisfied using Web Portal technology interfacing to Grid computing, data and information management and collaboration resources.

It will be used to guide the work done to establish a Sakai-based Virtual Research Environment for NCeSS and the ESRC e-Infrastructure.

Contents

1 Introduction

ESRC has recently established a small project to deliver a prototype e-Infrastructure for Social Scientists. This project makes use of collaboration tools and middleware which was developed in JISC VRE projects such as GROWL and the Sakai Demonstrator and NGS Portal to access NGS resources and data sources such as UKDA and MIMAS. It plans to link to data such as Census aggregate statistics via OGSA-DAI middleware and services being deployed in the projects such as GEMS by extending the GROWL tookit. Geospatial data is also involved, for instance in defining land boundaries for policy makers and appropriate tools have been developed at Leeds. The project heavily leverages previous experience and funding for related activities of the partners. It also has important consequences, for instance the outcomes of the project, which will also deliver a range of semantic tools, could be incorporated into existing JISC Information Environment (IE) services such as the IESR or IEMSR, cross search and repository services. We note the complexity of data requirements in this project has already been identified and documented by Miller [13].

This e-Infrastructure project does not currently have any provision for using such JISC IE services, such as other means of data and information discovery or publication of results – these are areas which could usefully be explored. A simple first step would be to include the SPP portlets in the Sakai portal to be deployed for the NCeSS hub and nodes, thus enabling interaction with SOSIG and other RDN (now Intute) and institutional repositories. Progress is being made on this with funding from CREE-2. Portlet interfaces to MIMAS and UKDA and other services useful to social scientists should be explored as should licensing of widely-used commercial software to run on the NGS and linking of remote datasets into such software. These services should offer an API (such as Web services) by which they could "plug in" to whatever portals or rich desktop client applications researchers decide are best for their projects.

Particularly in diverse and complex research fields like Social Science, personalisation and customisation is important. It helps the researcher to formulate questions and use the data and information available to support their new hypotheses. Data is subject to interpretation, the results of which need to be pushed into an information environment where they can be shared with colleagues, possibly providing alternative insight. An example would be an economist working with a child psychologist to interpret the effect of changes in school policy on educational attainment. It is somewhat strange that in the EVIE survey (see below) social scientists perceived the benefit of collaboration less than researchers in other domains.

For social scientists investigating worldwide trends, other sources of data and information outside the UK will also be required. This is particularly important as we seek to set up research partnerships, e.g. with NSF.

A simple all-embracing generic use case for "discovery to delivery" in research might be as follows:

```
A researcher wants to carry out a subject-specific search via one or
more portal interfaces and to be able to find relevant publications
and data associated with their studies and to be able to find other
papers which cite them. He/ she may also want to find associated grant
references and appropriate funding opportunities for related work.
```

```
The researcher then wants to access and download some of the datasets
and carry out a similar piece of work using a new model, new insight
or adding new data to the previous study. In an experimental study
they might be repeating a recommended procedure on one or more new
samples or applying an improved procedure to a benchmark sample.

The researcher will afterwards discuss and share results with a peer
group, using appropriate personal and group information management
software and will eventually create reports and publish the results
together with related data and model information.
```

More generally, we have found the key areas which need to be addressed are those of: integrating information and data; long-term archival and persistent access with appropriate access control; seamless search and discovery from a portal interface alongside other research tools; publication of data from personal and group information management systems; collaborative working in discovering, interpreting and using data and information. These broad areas, with subject-specific differences in detail and usage pattern, are constituants in the generic research life cycle and some aspects overlap with e-Learning and Digital Information management.

We consider the activities involved in doing research to be ultimately driven by knowledge creation. We make the following definitions (as presented to members of CURL in a meeting in October 2004):

Data: bits and bytes arising from an observation (non-repeatable), an experiment (repeatable) or a calculation;

Information: relationship between items of data of the form "A is always associated with B in some way".

Knowledge: understanding of causality in relationships "B happens after A because of X". This knowledge is shared globally.

In Figure 1 we show our own version of the research lifecycle steps which we believe to be appropriate to extended IE activities. This picture has been presented at NCeSS Steering Board meetings.

We have omitted from this the activities involved with grant proposal and funding, admin, collaboration forming, actual collaboration and computing as these are probably outwith the forseeable e-Infrastructure activities, but may be appropriate to other follow-on projects.

A full worked example scenario of a social science researcher engaged in quantitative research was given as input to JISC in [1] and is re-produced in Appendix B, this is again in the context of the JISC Information Environment.

It became clear during the study undertaken for JISC, that attempts to develop interoperability standards and provide single points of access to data and information are relatively immature. There are interesting issues at each stage of the process. We have taken into account end-user requirements, and could also analyse the process of publishing, discovering or accessing information. The Open Access community is starting to do this and are active in developing solutions to the various stages.

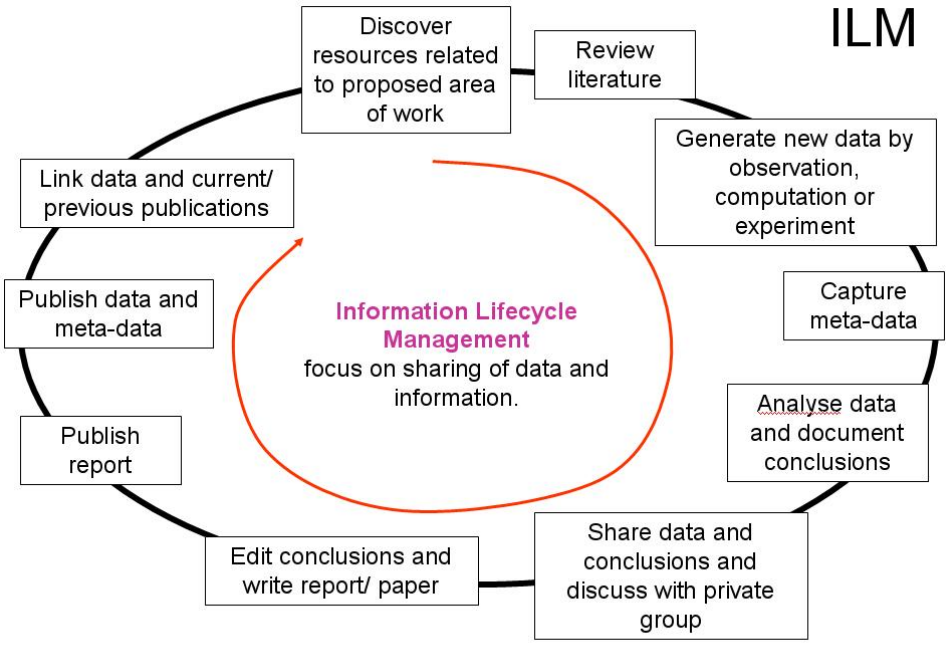

Figure 1: e-Research Data and Information Lifecycle Management

This embraces self archiving in relevant repositories (we have identified the use of institutional and facility repositories) and OA publishing (which introduces publishers' on-line repositories, e.g. domain-specific journals).

Before giving more general conclusions and recommendations, we pull together conclusions from the various supporting study areas which complement the e-Infrastructure Project.

2 Input to this Report

Input to this report was taken from the following sources:

1. From the Lancaster workshop on *e-Research, Portals and Digital Repositories Workshop* [5] 6-7/9/06;
2. From the Manchester workshop on *Portal Usability for e-Social Science* 8/11/07;
3. From scenarios, use cases and reference models;
4. From interviews, surveys and previous questionnaires;
5. From the survey of data sets D1.1.1.

These all addressed different aspects but provided insight to help our work.

2.1 From the Lancaster Digital Repositories workshop 6-7/9/06

At this workshop we presented the work we had done for the JISC on "Information Environment and Portals: meeting the Needs of Researchers". There were a number of related presentations and a discussion of wider issues focussing on the use of digital repositories of various kinds.

The main research resource discovery issues drawn out from this meeting include the following. There is a need to emphasise the relationship between data and outputs. Preservation and later access have to be addressed. Users need easy access to resources and tools to find resources. The perceived barriers include IPR, prestige, etc. and are not just technical. There have to be more resources for people to discover - should there be incentives to encourage more researchers to deposit data, results, reports and publications in repositories? Sharing data is considered important, but only in some disciplines. Could all benefit from more shared data? We need people to adopt and use existing services such as IESR and OpenDOAR as these could facilitate easier resource discovery.

There are many perceived gaps which include the following. There needs to be an awareness of resources and services available for research – researchers should check the library site first and consult with their information professionals. Where should we expect researchers to go for information? Is targetted advertising of research services appropriate, like Google or G-mail, e.g. a registry to pop up a suggested resource? As an example, the Blinkx Pico search engine, has equivalent of Google desktop. How does an organisation decide to target promotion of services to their researchers?

There is sill an IPR quagmire which includes licensing and usage – clearly articulated policies are needed. In a portal we need to be able to click a button to say "use this condition"? Need clarity of what can/ can't be done, remember that "open" does not mean completely sharable everywhere. NCeSS, JORUM, etc. are tackling some of these issues.

Funding councils can mandate deposition, but they should ensure that people understand what this means. RCs currrently rarely enforce or police their policies. ESRC has in some cases enforced deposition by withholding last grant instalment until it is done.

Resource discovery is mainly about finding publications and data, but we also need to find people. This could be via a partner-finding service, e.g. like EUservice. We are assuming that technology can solve all our problems, but actually there are human elements too – networking is a skill. How can we get researchers to use tools built on FOAF and RDF? Currently uptake is domain-specific, small groups might just use the phone. We need to understand *why* people are sharing information. By making information available, they might find new ideas from other people. The CONNECT portal was mentioned in this context.

As to current research, Web 2.0 is being widely considered to enhance the user experience. An outcome of the CREE project is that people would use resources if they are shown how to and its valid and useful for them to use. They will then explore at more depth. Web 2.0 can help to give an easy point of entry to explore further. A Firefox search box could include appropriate plugins. It however needs to have access to useful information or it will not be used. For instance, not all RSS feeds are kept up to date and people stop using them. Availability of information about resources requiring subscription is also desireable. This depends if metadata is publically available. Typically OpenURL resolvers will only find things to which which researchers are subscribed via their institution.

What facilitates working collaboratively and cooperating? We should demonstrate creating virtual

organisations – there is a perception that Shibboleth will help. Not all members of a VO will have the same access rights, e.g. because of institutional subscriptions. Annotation services may be important and there could be a separate annotation database, with attribution.

There are also sustainability issues: who will trust outcomes of a project which has finished, who will maintain outcomes? More "how to" information is needed to facilitate collaborative work to sustain open source and knowledge of using e-Research tools. How can we move development projects into service? Many projects are just survey and proof of concept. Typically a second phase moves them into a prototype. 3rd stream funding is required to realise a full service. Retaining skills and knowledge is clearly also important.

The final discussion points were on linking e-research, portals and digital repositories, as well as how to prioritise for the future. We need to get more academics interested in doing the development needed for specific the research areas. We need to consider services that are reacting to a needs of the community. One of the current disjoints, is how researchers handle their own research data versus how they use other people's data – tools to facilitate interoperability between public data and personal data could be useful. At the moment the user has to go to the publisher to get data or information not vice-versa, which means users have to know what exists – but currently many (most, >70% according to a recent core resource discovery study) don't know what the services are or what they can do with them. We also need to learn from failures – someone from outside a specific research domain will be confused by the alphabet soup and unable to collaborate. Perhaps a study that analyses what a research domain has done and what's available, e.g. in the form of a flow chart could be useful. Research communities are not inclusive, i.e. people who've been funded don't know one another, and are not sure about the links between projects. How is the success of a development project judged? Who does the evaluation and why? What are the metrics?

2.2 From the Manchester Portal Usability workshop 8/11/07

A Sakai server was set up and customised to have the same look-and-feel as the NCeSS Web site. This is running at `http://portal.ncess.ac.uk`. A worksite in this portal was set up for attendees at the Portal Usability Workshop, and populated with a few of the more commonly-used tools: Resources; Blog; Wiki; Chat; Schedule; Announcements; Discussion; News; Site info. During the workshop other tools were added as their capabilities were discussed, including: Search; WSRP; Agora; Polls; Forums; Forms. Agora was demonstrated from a local server as was the Portal Access Grid, although these demonstrations were somewhat limited by the wireless network (no cable access was available in the building).

The Sakai Wiki has been used successfully to write collaborative articles (used in a VRE-1 project for the Early Modern Virtual Research Group). This was demonstrated. A Yahoo! Map mashup tool developed for the Sakai Demonstrator was also shown. The portlets being developed in the MoSeS project at Leeds were also demonstrated.

Participants experimented with configuring and using the tools, adding new users, changing permissions, etc. Comments from the workshop were recorded by notes (thanks to Grace de la Flor and Derek Sergeant) and on-line using the Wiki, Blog, Discussion and Forum tools.

Some of the comments, anonymised and in no particular order, are as follows:

- Announcement tool
  - allow users to filter by announcements that have already been read, allow users to delete ones they no longer need
- Calendar tool
  - needs to be interoperable with formats such as; Exchange Server, vCal and iCal
- Discussion tool
  - is for threaded discussions, but the buttons are non-intuitive. For instance, would have expected a "post" button.
  - got stuck in an un-defined state.
- Chat tool
  - Sakai needs to show all the users that are on-line
  - appeared to hang with a lot of simultaneous users, or at least it was very slow
  - Can you get the chat tool to scroll to the most recent part of the chat by default when you follow the link?
  - Wouldn't it make more sense to put new comments at the top and push the old ones down? Otherwise, coming in will always show you the oldest stuff by default, rather than the newest.
  - It really should at least scroll down (needs javascript usually)
  - wouldn't it make more sense for the chat to automatically scroll to the bottom when you select the 'chat' link? I was away blogging then came back here and if you had a really long chat then scrolling down all the time might be a pain
  - A search on a name doesn't show chat entries made by that person (but shows entries that mention them)
  - I think there that to post messages without needing to refresh you need to click on "add message" rather than pressing return.
  - Chat posts sometimes appear duplicated, but if you refresh the browser window then the duplicates disappear
  - Some messages appear only when the page is refreshed (I think I saw this in the JIRA somewhere, so may be a well-known bug).
  - The user list also does not get updated in the background.
  - the recent chats list that displays on the home page does not contain any links – this means you have to use the lefthand menu to go there
  - I have also seen messages scrolling down below the input box
- Blog tool
  - how easy is it to find Blog posts? Not that easy – for instance, you can't tell which posts have been commented on!
  - nice feature that you can delete older posts, even if you have to be a maintainer to have deletion rights

  - Clicking the box doesn't show my cursor until I start typing. (Also true in the abstract box).
  - seems really a bit too complicated, given what it is trying to acheive. My first try, I hit "Add to Document" and then left the tool. This appeared to loose my entry with no warning that it would be lost. This is because the "Save" button was not visible on my screen at this point!
  - The "keywords" box also seems odd - it starts filled with text which disappears when you type. It doesn't highlight like normal. It would be better to put the help text outside the text box (this would also help if you forget for instance that you need commas to separate the keywords). It may also be useful to have a list of existing keyworks to select from.
  - pictures in the Blog do not get re-sized. This is the same in the Wiki.
  - can the Blog be made public like the Wiki?

- Worksite setup tool
  - The worksite setup tool has buttons that allow you to skip through the various worksites but doing this does not change the active worksite. This can lead to confusion when people look at the tab bar to figure out what it is they are editing.
  - We suggested we could remove the "next" buttons from this tool to avoid accidents.
  - When using the worksite configuration tool on "my workspace" and editing tools, the list of tools selectable is quite limited (no blogging, for example) and there is no indication as to why that is the case.
  - Another problem is that for some of the tools currently selected there does not seem to be a tick-box so that they can be deselected?

- Search tool
  - the search function does not search polls
  - does the search content search indexed content or just metadata?

- Wiki tool
  - Can Wiki pages could be made public? It turns out that all that is needed is to tick the public access box in the Wiki page info and point people to the public view (under feeds, other feeds are accessible as well). So, create a link from somewhere, done. Wiki links get translated appropriately, i.e., where a publicly accessible page is linked to the URL for the public view is used and there a page is private no link is generated.
  - Of course, all resources used in the Wiki pages have to be public as well.
  - One feature request for the Wiki would be to have default settings or a hierarchical access control system (but perhaps I just haven't found this yet).
  - Wiki 'edit' should read 'add/edit page'
  - Needs help page for wiki syntax and/or allow traditional html posts
  - 'Page does not exist' should read 'Add content'
  - From Ian Boston in response to some of these comments: in general if you want something fixed or changed the fastest route is to create a patch, describe what you are trying to do in a Jira, attach tha patch and assign to me (for Wiki). I apply patches faster than I fix bugs :)

- General Usability Issues
  - Disparity between tools, e.g. Discussion posts have reply counts, but Blogs comments don't. It would be good if the Blog could display when replies are posted. Why isn't there a recent Blog or Discussion activity on the home page like there is for the Chat?
  - Some of the ways in which Sakai maintains state are less than intuitive. It is possible, for example to go to "my workspace", select the "worksite setup" page and end up editing the worksite setup for the last worksite edited rather than the one just selected!
  - Notice that when you create a new user account as a normal user, you will get a "Site unavailable" page displayed. This won't affect the newly created account, i.e., when the new user log on, he/she would see his/her "My Workspace" worksite.
  - Content/ service providers need to provide Web access to their databases
  - Portlets and their APIs need to be open source
  - Sakai is a "collaborative learning environment" providing "rich desktop services" The tools available have a history in the educational support of teaching and learning It offers general collaborative tools
  - Developers need info on how to integrate their portlets into the portal
  - it would be quite nifty to have all the tools in loose little frames, like Agora, so you can chat while you are video conferencing in one overview
- The framework doesn't impose constraints for the e-Social Science portal
- Questions about migration to future Sakai releases, e.g. Sakai 2.5 has some potentially interesting features such as iCal import and export
- Mashup technology described in a recent talk by Chuck Severance could be of interest
- Possible new tools for the portal could include Connotea, reference management, Open social-like applications, Google maps

2.3 From the scenarios, use cases and reference models analysis

We note that a set of case studies has been carried out by the Oxford e-Science Cente, `http://www.ncess.ac.uk/research/social_shaping/oess/in_depth/`. The 5 in-depth case studies looked at projects: MiMeG; Digital Records; eDiaMoND; Advanced Grid Interfaces; and the Reality Grid.

The Oxford team however noted: *The design and use of advanced Internet and Grid tools and infrastructure in the social, physical and computer sciences are likely to re-configure not only how researchers produce, use and collaborate around key resources, such as data and software tools, but also how they share such resources and who can gain access to them. This reconfiguration of access raises numerous ethical concerns, legal uncertainties, and institutional conflicts. To explore these interrelated phenomena, we have re-designed our in-depth case studies to move away from a "project" focus to aim instead at the study of data sets and tool artefacts with a focus on the practices and issues that emerge as the data and tools are developed and travel between individuals, groups, communities and different stakeholders.*

The following set of conclusions arose from our analysis of a number of previously documented use cases and scenarios from relevant projects [1].

- Researchers want access to data and information (e.g. scholarly publications) for a variety of reasons. They want to access all sources in a seamless way and to have a uniform style of presentation;
- They want to use the results of such discovery for a variety of purposes, fusing data and information from multiple sources;
- They want to use previously stored data and also create new data and information from computational or experimental procedures;
- They want to publish new data and information, potentially from personal repositories into public repositories;
- Research Reference Models can be developed based on research processes outlined in the scenarios and use cases;
- these RRMs represent parts of the generic Research Lifecycle;
- RRMs can be realised as Designs using generic service components (this hypothesis is yet to be fully tested);
- The IE Architecture can be extended with additional components to accomodate an implementation of these designs in real artefacts [1];
- A range of context-based user interfaces are required to access components in the extended IE architecture;
- Use of the components and services can be facilitated by workflows supporting the research process;
- Many activities worldwide are beginning to implement parts of this overall architecture and we need to integrate with them. This includes information and data services developed in e-Science programmes in USA, Europe and Asia-Pacific regions;
- However, toolkits to support the implementation of most of the client-side services are not yet available and portals currently provide a usable option for Web-based access.

2.4 From the interviews, surveys and previous questionnaires

These conclusions arose from our analysis of a number of previous surveys and questionnaires supplemented with interviews with key stakeholders [2].

Linking research practice, resource discovery and information retrieval needs an environment into which they are all integrated. We found that the previous surveys have taken too narrow a view of this, since they have mostly been discipline specific or have focussed on one aspect of this activity. The joint space requirements still need further investigation, i.e. computing and collaboration, or personal information management and admin functions. Portals tend to provide a set of customisable

[1] this was in the language of the original international e-Framework for Education and Research. Currently they are loosely referred to as "technical components". Reference Models are currently referred to as "Service Usage Models". See `http://www.e-framework.org`.

but pre-defined tools and services which are less flexible than a rich desktop application, although the development of interface technologies such as Ajax, and those suggested by the FLUID project are modifying this perspective in interesting directions.

However, with regard to research, some key conclusions can be drawn out from previous studies, including that:

- Researchers need access to data storage and computational resources, as well as software and services;
- Provenance is key to establishing the quality, reliability, and value of data in the discovery process (this has also been noted by the DCC);
- Any interface needs to present the views of multiple services in a way that is easy for users and administrators to access and customise;
- There is a need to understand more fully disciplinary differences in user requirements (we looked more at this in [1]. Research issues which have been raised in previous studies are related to data format diversity as well as meta-data, mapping and vocabulary;
- Existing services and methodologies could be shared and Web-based presentation layers customised for delivery to users, e.g. in portals;
- A range of toolkits (thin clients, portals, scripting languages, GUIs etc.) should be developed to extend and simplify access to Grid resources and information systems leading to the eventual emergence of one or more interfaces to a Virtual Research and Information Environment. However this requires the existence of a set of underlying re-usable services. Any such services which have arisen from the e-Science Programme and JISC VRE programme are currently very domain-specific and require expert knowledge to use.

With the development of e-Research groups, new needs appear to have emerged. It is likely that the needs that will be important for a given institution will vary by the:

- Areas of research strength;
- Extent of infrastructural development;
- Strength of collaborative networks.

It is clear from the surveys of user requirements that researchers need access to scientific and other data as well as publications. Whilst it is probably not within the remit of the project to host such data, it may be appropriate to consider how corresponding meta-data might be hosted or to provide search facilities and mechanisms to link data to publications and *vice versa*, perhaps working together with JISC.

Overall, researchers appear to need more support for learning, adapting, and writing software specific to their research problems than is currently available. Also, researchers who are generating and using large data sets need help managing their data. This need will become more pressing as data enters long-lived data repositories and therefore the public arena through preservation rather than simple publication of links on the Web.

2.5 From the survey of data sets

We suggest that the e-Infrastructure Project might usefully consider linking into a wider range of data services than envisaged at the start. These include private and commercial as well as open services. This view is confirmed in the report D1.1.1 [13], which documents a recent study performed to select datasets of primary interest. This illustrates the complexity of requirements from which we cite:

- British Household Panel Survey
- Census 1991 SARs (Samples of Anonymised Records)
- EDINA UKBorders Census boundary data for SARs 1991
- Health Survey for England
- National Child Development Survey
- Datasets from MRC and NERC as required
- Quarterly Labour Force Survey
- General Household Survey
- British Social Attitudes Survey
- Workplace Employee Relations Survey
- ONS Omnibus Survey
- ONS Millenium Cohort Survey
- International Monetary Fund (IMF)
- World Bank and Organisation for Economic Cooperation and Development (OECD)
- ESDS International service
- British Crime Survey
- ONS Neighbourhood Statistics
- European Social Survey
- Eurobarometer data series
- Administrative, retail, consumer, video CCTV and Web usage data

These datasets are from diverse sources, local, national, international; open, commercial and confidential. The same survey mentioned the following requirements related to access and tools.

- Shibboleth enabling
- GIS system to utilise boundary data
- Longer term access to data via Grid technology direct from provider
- Software and methodology from existing projects
- Access to datasets from other disciplines
- Tools for geographic mappings
- Meta-Data registries
- Controlled vocabularis and ontologies
- Question banks
- Classification schema and variable mappings
- Linking between data and related publications
- Make commercial tools such as SPSS, SAS, Stata available for Grid work
- Virtual safe setting for analysis of confidential data, e.g. Census Controlled Access Microdata Samples.

Whilst this is illustrative of the broad field of social science, similar data requirements and diversity of sources is typical of many research fields.

3 Input from other VRE Projects

We assume that VRE functionality will most often be delivered via a Web portal, perhaps through tools in a Science Gateway or Institutional Portal. We note however that current portals only address parts of the use cases. We here identify the broad requirements for resource discovery and portals within the e-Research community.

In writing this section we have also drawn upon discussions with developers and users in the VRE Programme and the user requirements studies they have carried out during and prior to this study [2]. We have also participated in workshops on usability and requirements, for instance the *Science Gateways* workshop at NeSC 19/5/06.

A typical research portal might involve effort from staff across the university or collaboratory and might provide seamless access to:

- My Research profile;
- Data warehouses that deliver business intelligence on research applications, awards and income;
- Costing and project management tools;
- Research publication databases and research expertise systems;
- Peer review tools;
- Library catalogues, bibliographic research resources and digital repositories;
- Access to shared facilities on the Grid and to primary research data and meta-data associated with relevant projects
- Service portals provided by Research Councils, government departments, etc.;
- Asynchronous communications – Email/ discussion fora;
- Synchronous communications – Chat/ shared whiteboard;
- Desktop video conferencing;
- Calendaring and meeting management;
- News – BLOG/ RSS feeds;
- Collaborative writing – Wiki;
- In the UK the RAE tool and the RCUK Je-S online application process.

As a further illustration, the functionality of Google and other major internet search engines has been illustrated.

In our consideration of deep search and discovery services and in reference to Google, we note that Web browsers are only one client for research services and that others, such as Web Services which

can be linked into "heritage" applications, GUIs etc. are likely to be of increasing future importance. Andy Powell [15] notes the potential of using RSS, iPod or Firefox plugins. We [18] have shown that this is possible using pattern-based Java technologies such as J2EE. End users are likely to require a variety of client tools for both machine- and human-oriented access including ones which can be used for management of their personal information. With the growing use of Web 2.0 technologies such as Ajax and mashups this need becomes even greater.

It is arguable that this is where e-Research technology can make the biggest impact and provide functionality through active links to a wide range of resources *not* simply accesssible from a Web browser.

Some more specific requirements for VREs that impact on the provision of repository services and personal information systems came from the Sakai VRE Project are [9]:

- Access to best-practice documentation, and support for best practices, within the VRE;
- Capture and storing of collaborative discussions;
- support in training new researchers
- Searchable list of conferences, lectures and other events;
- Locate other researchers;
- Selective delivery of information;
- Supporting grant applications;
- Forums and spaces for internal communication and recruitment;
- Access to searchable databases of digital (digitized) artefacts;
- Data repositories.

And another set from the EVIE VRE project are [16]:

- Find and acquire published information such as articles, conference proceedings, literature;
- Find out about funding opportunities; apply for funding; managed funding projects;
- Collaboration with partners with the University or at other institutions ;
- Share or archive research results such as preprints, postprints, technical reports, software, or datasets;
- Other activities.

Questions leading to these responses had been asked in terms of a research life cycle similar to the one outlined above.

Every faculty rated the activities surrounding resource discovery as the most important for a VRE to support, with 70% of respondents rating it as essential.

Funding opportunity tasks are also rated as very important, with some faculties rating it as having the same importance as resource discovery but the Faculty of Medicine and Health and the Faculty of Education, Social Science and Law indicated that these tasks need not be supported as strongly for their disciplines. One comment suggested that there are several information sources about funding opportunities already available together with support and advice networks, so this provision might fall outside of the VRE.

Collaboration activities were rated as very important or essential by over half of the respondents but this is very uneven across the faculties. The Faculty of Arts and the Faculty of Medicine and Health predominantly rated this aspect of the research lifecycle as important or somewhat important. It was surprising that the activities surrounding managing research outputs received this low rating, as the one-to-one sessions had suggested more interest. Also, this area of the research lifecycle was seen as not important by 12% of respondents. These "not important" responses came entirely from the following five faculties: Arts; Performance, Visual Arts and Communication; Business; Education, Social Science and Law; and Biological Sciences. For these faculties more respondents rated managing research outputs as not important than as essential.

At this point in the survey it would not have been obvious which activities might come under the catch-all aspect of other activities. This meant that this area of the lifecycle was only rated by half of the respondents.

The distribution of the five importance ratings across the aspects, when broken down by the research level of the respondent, is proportionately representative of the overall ratings, with just one exception. Only 10% of graduate students and post-doctoral researchers rated the funding opportunity activities as essential, where, overall, 30% of respondents rated this aspect as essential. This is attributable to most graduate students having no interaction with funding applications.

This report also identified priorities in terms of portal functionality and usability.

A survey of the kind of digital library services currently used by researchers is presented in a separate document [3].

3.1 DAMES: Data Management through e-Science

The DAMES node of NCeSS (funded in 2008) aims to deliver three linked portal interfaces: GEODE for occupational data; GEEDE for educational data; and GEMDE for ethnicity and migration data. These will link into service providers CESDA-PPP and UKDA/ ESDS. The portlet interfaces are being developed using GridSphere, but will ultimately be hosted by the e-Infrastructure Sakai VRE. The project also addresses generic services for cross-searching of data sources using appropriate metadata, and linking to analysis software such as Stata and SPSS. Clearly authorisation is an important area in this project for access to data and licensed software and Shibboleth will be used as in the SARoNGS project.

3.2 MoSeS: Modelling and Simulation for e-Social Science

MoSeS focusses on development of a national demographic model and simulation of the UK population specified at the level of individuals and households. MoSeS will help town planners to forecast trends in healthcare, business and transport in policy making by predicting demographic changes looking forward up to 20 years.

To make predictions, computationally intensive agent-based simulation models are run on the NGS using a number of distributed data sources. MoSeS links into services such as the Census for current and historical demographic information and the EDINA geo-linking service for mapping data on different regions such as electoral districts known as wards.

MoSeS offers users a number of different ways to interact with its simulations and scenarios. A series of JSR 168 portlets have been developed for the MoSeS portal, which uses the GridSphere framework. The same portlets will work in the e-Infrastructure Sakai VRE. From this interface, PDF reports of simulation results can be generated via a simple workflow. These reports contain maps, tables, comparisons, etc. based on user-defined criteria. It is also possible to visualise simulation results using the Google Maps service or stream results into Google Earth.

3.3 PolicyGrid, GeoVue, Obesity e-Lab, GENsSIS

In addition to Sakai and GridSphere, several NCeSS nodes have their own Web 2.0 style applications. These include ourSpaces from Policy Grid (Aberdeen), MapTube from GeoVue (UCL), and MyExperiment used in the Obesity e-Lab (Manchester, funded in 2008). The OurSpaces community Web2.0 portal currently has functions for Search; Upload; Squanto; EviCon; SPSS analysis; FEARLUS simulation.

GENeSIS (Leeds and USL, funded 2008) has demonstrated using SecondLife as a on-line urban laboratory to explore issues pertaining to planning and public debate in a virtually 3D collaborative environment. It is difficult to integrate such self-contained Web applications with a portlet framework unless their services can be exposed using an appropriate API such as JSR 168 or WSRP and data-interchange standards. We nevertheless concede that there is unlikely to be one single VRE that meets all requirements and a traditional portal may not always be the first choice.

4 So, what Functionality can we offer Social Science Researchers?

We have taken into account the various inputs summarised above, and evaluated the capabilities of a number of portal frameworks. From the latter we concluded early on that the Sakai framework, with its origins as a collaborative learning environment, could be easily adapted to support the needs of e-Research. Following the work of several JISC VRE projects this is today even more true.

A portal for e-Research is likely to require the following, though we note that this list still needs further expansion testing with users.

- Information and data from institutional and external sources:

  - Access to full text resources as well as tools to cross search both free and subscription based services. The results pages should identify subscription based resources and whether the user can gain access to them;
  - Access to departmental and local resources and repositories as well as external resources from one interface;

- Collaboration and research resources:
  - A mechanism to provision members (people, devices) in collaborative sessions. To include shared access to data repositories for searching, replication and updating. This requires ID and access management across institutions;
  - Applications such as Web pages, shared presentations – in an environment like Access Grid these are driven by a presenter from a master document and can also be viewed in a portal version;
  - Generic tools: text chat, white boards - need shared updates to text message streams;
  - Audio-video conferencing and collaboration tools – to share events specifying changes in compressed streams;
  - Visualisation – to share events corresponding to changes in pixels of a frame buffer, maybe using SVG;
  - Shared maps, instruments, (e.g. medical);

- Training resources:
  - Alert services, promotion and training opportunities on how to use services accessed from the portal.

Any provision also needs to be supported by assistance with organising and managing research data sets, as well as a training programme.

The Table 1 lists the capabilities of Sakai tools in terms of "out of the box", "implemented in VRE programme", "available elsewhere as portlets", "desireable".

We have omitted e-Learning tools from this list, but they could play a role in the training requirement. We have also not discussed in this report issues relating to the organisation of the content of a portal, e.g. relating to the definition of VOs, groups and access rights which can be realised via Sakai worksites and roles.

To date, we have deployed a Sakai-based portal for NCeSS and the e-Infrastructure project which is available at `http://portal.ncess.ac.uk`. This is being populated with tools and capabilities as outlined above based on discussions with project partners. This work is also being guided by the discussions at the Portal Usability Workhop and with other e-Researchers in the UK and elsewhere. A view of the home page of the e-IP Project worksite is shown in Figure 2.

What is currently missing, except for the Leeds suggestion to add their GeoLinking Service, is an understanding of the research-specific tools required by the NCeSS nodes and related projects. This is perhaps a question that the Hub could help address.

Table 1: Sakai Tools for e-Research

Out of the box	Implemented in VRE Programme	Available elsewhere as portlets	Desireable
Home	Agora	Portal Access Grid [1]	Connotea style reference management
Announcements	Conference Document Management	GeoLinking Service [2]	Google maps
Blogger	WSRP Consumer	Grid JSDL portlet [1]	open social-networking apps
Calendar	Wiki	MyProxy Manager [3]	Google gadgets
Chat Room	Shibboleth login	Grid file transfer [3]	
Discussion	Portlet Bridge	JAFER search [5]	
Email Archive	Semantic search	Google search [5]	
Forums	Condor (Campus Grid)	Bookmarks [4]	
Glossary	Map Mashup	Issue/ Bug tracker [4]	
Messages	UDDI Registry interface	MS Exchange [4]	
RSS News		Amazon search [4]	
Podcasts		JNDI [4]	
Polls		Notepad [4]	
Resources		Alfresco CMS [4]	
Search		Project Management [4]	
Site info			
Tests and Quizzes			
Web Content			

1. OMII portal project
2. from JISC OGC Grid Collision SeeGeo project
3. NGS Portal
4. see eReSS Wiki `http://penfold.lib.hull.ac.uk:8080/confluence/display/context/Portlet+Registry`
5. CREE and CREE-2

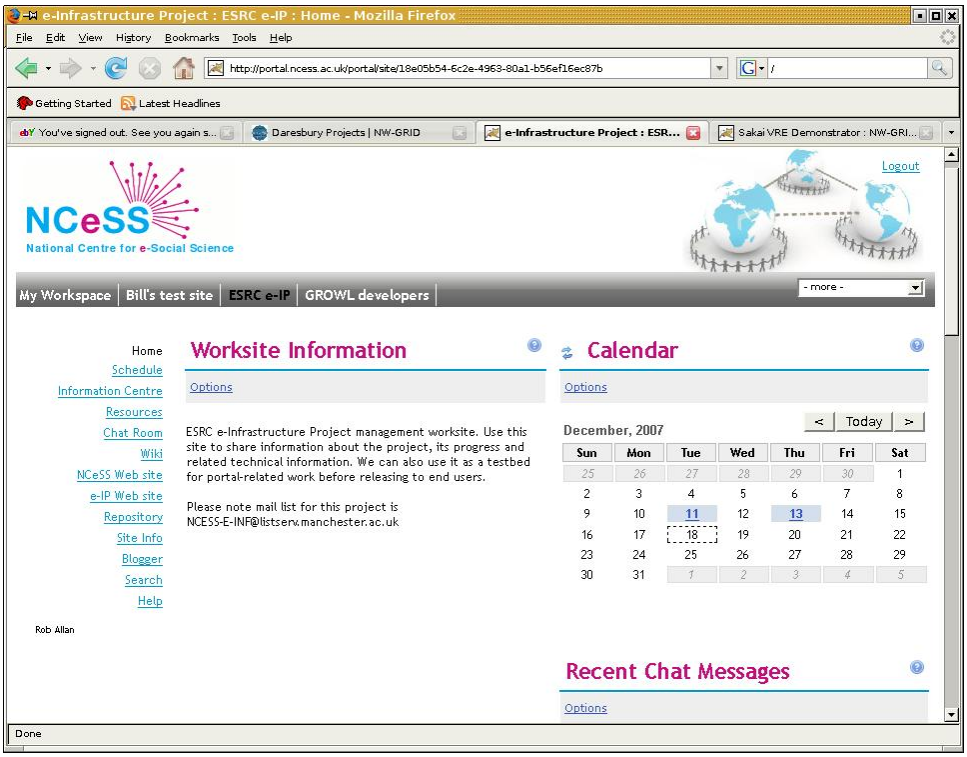

Figure 2: NCeSS Portal

5 Acknowledgments

ESRC, JISC and NWDA (via NW-GRID) for funding.

We thank people we spoke to: Ties van Ark, Chris Awre, Dexter Canoy, Keith Cole, Ian Dolphin, Matthew Dovey, Pete Edwards, Pascal Ekin, Adrian Fish, Grace de la Flor, Carole Goble, Chris Higson, Caroline Ingram, Jiten Baghat, Yuwei Lin, David Meredith, Dave de Roure, James Reid, Derek Sergeant, Robert Sherratt, Rich Sinnott, Andrew Smith, Steve Swinsburg, Frederique van Till, Paul Townend, Andy Turner, Martin Turner, Alex Voss.

References

[1] R.J. Allan, R. Crouchley and C. Ingram *JISC IE Portal Activity: Scenarios, Use Cases and Reference Models* (CCLRC, June 2006)

[2] R.J. Allan, R. Crouchley and C. Ingram *JISC IE Portal Activity: Comparison of Surveys* (CSI Consultancy, June 2006)

[3] R.J. Allan, R. Crouchley and C. Ingram *JISC IE Portal Activity: Web-based Library and Information Services* (CCLRC, June 2006)

[4] R.J. Allan, R. Crouchley and C. Ingram *JISC IE Portal Activity: The Information Environment and e-Research Portals* (CCLRC, June 2006)

[5] R.J. Allan, R. Crouchley and C. Ingram *e-Research, Portals and Digital Repositories Workshop* Notes from the workshop held at University of Lancaster 6-7/9/06 (CSI Consultancy, September 2006)

[6] R.J. Allan, R. Crouchley and C. Ingram *JISC IE Portal Activity: Final Report* (CCLRC, June 2006)

[7] Christine Borgman *Building a Usable Infrastructure for e-Science: An Information Perspective* Keynote talk at the UK e-Science All Hands Meeting 2005, (Nottingham, UK, 19-23 Sep 2005) `http://www.nesc.ac.uk/talks/ahm2005/keynote1.ppt`

[8] Catherine Jones et al. *CLADDIER: Citation, Location And Deposition in Discipline and Institutional Repositories* `http://www.jisc.ac.uk/index.cfm?name=project_claddier`

[9] Graham Klyne et al. *Sakai VRE Demonstrator Project User Requirements* OSSWatch Wiki (2005) `http://wiki.oss-watch.ac.uk/SakaiVre/UserRequirements`

[10] Liz Lyon and Andy Powell *JISC Information Environment (IE) Inventory with Diagrams* (UKOLN, 2005) `http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/`

[11] Matthew Mascord, Marina Jirotka and Clint Sieunarine *Integrative Biology VRE, Work Package 2: Initial Analysis Report* `http://www.vre.ox.ac.uk/ibvre/IBVREInitialAnalysisReport.pdf` University of Oxford (November 2005)

[12] Ma Moraga, Coral Calero, Mario Piattini, Oscar Diaz, *Improving a Portlet Usability Model* Software Quality Journal, Volume 15, Number 2, (Springer, June 2007) pp155-177(23)

[13] Ken Miller *Primary Selection of Datasets* Deliverable D1.1.1 of the ESRC e-Infrastructure Project (August 2007) `http://www.ncess.ac.uk/services/research`

[14] Andy Powell *A 5-step Guide to becoming a Content Provider in the JISC Information Environment* Ariadne 33 (2002) `http://www.ariadne.ac.uk/issue33/info-environment/intro.html`

[15] Andy Powell *The JISC Resource Discovery Landscape* Report to the JISC (UKOLN, May 2005)

[16] D.M. Sergeant, S. Andrews and A. Farquhar *Embedding a VRE in an Institutional Environment (EVIE). Workpackage 2: User Requirements Analysis* User Requirements Analysis Report (University of Leeds, 2006)

[17] M.J. Smith *Use Case Compendium of Derived Geospatial Data* (GRADE Project, December 2005) `http://www.edina.ac.uk/projects/grade/usecasecompendium.pdf`

[18] X. Yang, A. Akram and R.J. Allan *Developing Portal/ portlets using Enterprise Java Beans for Grid Users* Concurrency and Comutation: Pract. Exper. Special issue on Portals. 19 (2007) doi:10.1002/cpe.1069

[19] *DigiRep Wiki* the on-line Wiki resource for the JISC Digital Repositories Programme. It is maintained by UKOLN `http://www.ukoln.ac.uk/repositories/digirep/index/JISC_Digital_Repository_Wiki`

[20] *NGS: the UK National Grid Service* http://www.grid-support.ac.uk

[21] *JISC Virtual Research Environments Programme* http://www.jisc.ac.uk/index.cfm?name=programme_vre

A Glossary, Abbreviations and URLs.

A glossary with many relevant entries can be found at: http://www.grids.ac.uk/ReDRESS/glossary_v2/glossary_v2.html.

Wikipedia can be used to obtain an explanation for most of the generic ones, http://www.wikipedia.org.

Specific abbreviations used in this report are:

CCLRC: Council for the Central Laboratory of the Research Councils, now part of STFC

Condor: Middleware for Campus Grids. Includes DAGMan which is a workflow management tool.

CONNECT: The Higher Education Academy On-line Portal http://www.connect.ac.uk

CREE: Contextual Resource Evaluation Environment, JISC-funded project http://www.hull.ac.uk/cree

CSCW: Computer Suppoerted Collaborative Working, see Wikipedia

CURL: CURL: Consortium of University Research Librarians

DCC: Digital Curation Centre http://www.dcc.ac.uk

DRM: Digital Rights Management, see Wikipedia

ePubs: STFC ePubs open access repository http://epubs.cclrc.ac.uk

CQeSS: Collaboratory for Quantitative e-Social Science (NCeSS Node)

eReSS: JISC e-Research Interoperability and Standards project

ESRC: Economic and Social Research Council http://www.esrc.ac.uk

GEMS: Grid Enabled MIMAS Service, JISC-funded project http://pascal.mvc.mcc.ac.uk:9080/gems

Grid: A collection of distributed computing and data resources connected by middleware.

GROWL: Grid Resources on Workstation Library, JISC-funded VRE-1 project http://www.growl.org.uk

HCI: Human Computer Interface, see Wikipedia

IEMSR: Information Environment Meta-Data Schema Registry http://iemsr.ac.uk

IESR: Information Environment Service Registry `http://iesr.ac.uk`

Jini: Java services based middleware

JISCmail: JISC mail service `http://www.jiscmail.ac.uk`

JXTA: Java peer-to-peer middleware

MOSeS: Modelling and Simulation for e-Social Science (NCeSS Node)

NCeSS: National Centre for e-Social Science `http://www.ncess.ac.uk`

NGS: National Grid Service `http://www.ngs.ac.uk`

OGSA: Open Grid Services Architecture

OGSA-DAI: Data Access and Integration using Open Grid Services `http://www.ogsadai.org.uk`

ONS: Office for National Statistics `http://www.statistics.gov.uk`

P2P: Peer-to-peer, see Wikipedia

RCUK: Research Councils UK `http://www.rcuk.ac.uk`

ReDReSS: Resource Discovery for Researchers in e-Social Science

RMCS: Remote MyCondor Submit workflow system from eMinerals

SOA: Service Oriented Architecture

SOAP: Originally stood for Simple Object Access Protocol, the basis of Web services

RRM: Research Referemce Models, now referred to as SUMs in the e-Framework, `http://www.grids.ac.uk/Papers/Classes/classes.html`

Sakai: Sakai collaborative learning framework adapted for research purposes `http://www.grids.ac.uk/Sakai`

SOSIG: Social Science Information Gateway, now Intute Social Sciences `http://www.intute.ac.uk/socialsciences/`

SPP: Subject Portal Project `http://www.portal.ac.uk/spp`

SRB: Storage Resource Broker `http://www.npaci.edu/DICE/SRB/`

STFC: Science and Technology Facilities Council, formed by combining CCLRC and PPARC, see `http://www.stfc.ac.uk`

VO: Virtual Organisation

WSDL: Web Service Desciption Language

WSRF: Web Service Resource Framework

Web Services (WS-I): Language agnostic remote method invocation using XML, SOAP, WSDL and UDDI

SUM: Service Usage Models as defined in the e-Framework for Education and Research, see `http://www.e-framework.org`

SVG: Scalable Vector Graphics, see Wikipedia

UDDI: Universal Description, Discovery and Integration, a Web services registry specification, see Wikipedia

UKDA: UK Data Archive `http://www.data-archive.ac.uk`

B Scenario from Social Science Research

A scenario is a short story that describes the functions in a context. To illustrate this we do not need need use cases as this is far too detailed for the broad picture we want to present. We show below how a scenario might inform the necessary components of an architecture for information and research services based on that of the JISC Informatio Environment. Similar analysis can be done for the other scenarios in Section 2.3. This set chould be widened based on other reviews and studies which have been completed recently. Our worked example is based on e-Social Science and its differences with other disciplines have been described by Borgman [7].

A possible Social Science Researcher's (SSR) scenario is.

1. Suppose we have a researcher (SSR) who could access all of the Archived Data Sets and those used in every social research publication in their research field and decide on the most appropriate data for their needs, without having to spend days reading through coding schedules and questionnaires;
2. Suppose SSR could automatically re-estimate all the models others have used on these data sets, and see what happens if you drop or add new variables to the analysis;
3. Suppose SSR could quickly formulate (check the identification etc.) and estimate any new models or combinations of existing models you thought might be relevant;
4. Suppose SSR could re-do this across multiple datasets;
5. Suppose SSR could match your research questions to information held in existing digital resources. Search for new explanations;
6. Suppose SSR could integrate multiple sources of data and text to help to fill in missing data and ideas.

Services (or steps) required in this scenario include:

- search publications and archived data sets;
- select and download appropriate data matching a particular research need;
- re-construct previously used models;

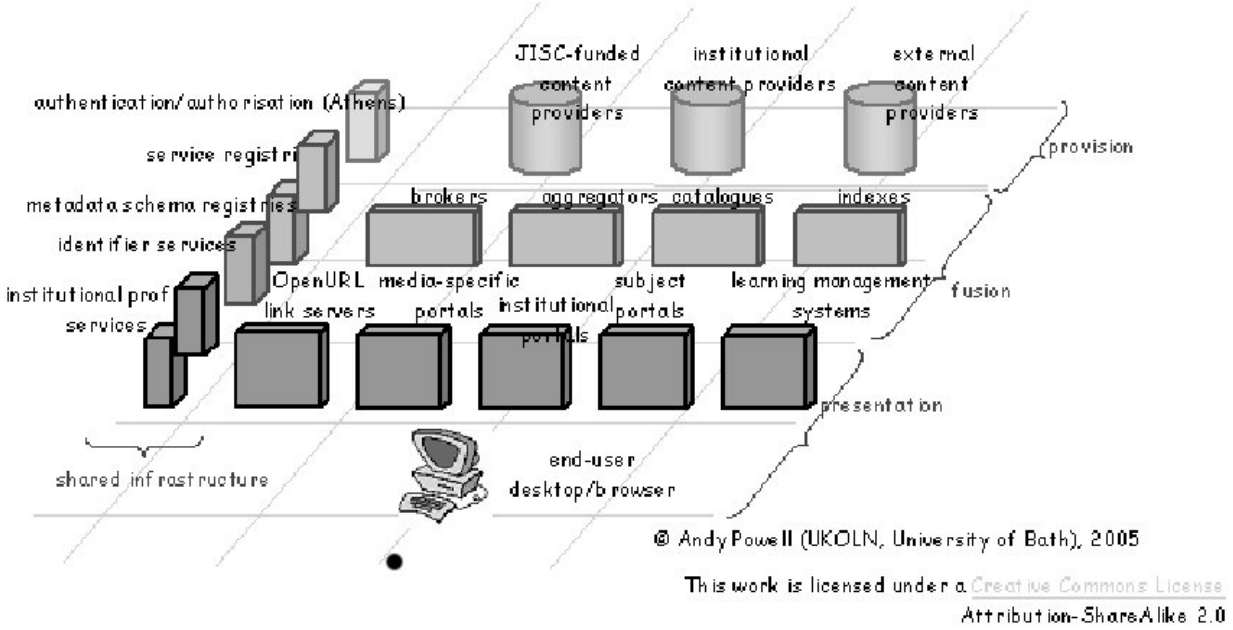

Figure 3: IE Architecture version 1

- re-compute models on these data sets;
- re-compute models on these data sets with different parameter choices;
- compare results;
- create new models or combine existing ones;
- repeat analysis across multiple datasets;
- match research questions to digitally-stored information;
- integrate multiple data and text sources to identify missing data and ideas.

What does this imply for the architecture? Well the original IE architecture diagram from Andy Powell [10], Figure 3 is missing a few key functions/ elements which should be present in an e-Infrastructure.

To illustrate the missing elements we can look at the mapping between the various suppositions of this scenario and the elements of the Powell's diagram.

1. At first inspection the notion of content as used by Powell and archived data sets may be different, so we need to be clear that MIMAS, EDINA and the various archives, ESDS data archive as well as the data archives of online journals that contain copies of the data sets used by journal authors, etc. So perhaps instead of content we should use the phrase Digital Repository (DR) in our use of these diagrams and make it explicit. Also its not clear what is the presentation layer that SSR would use. It could be a project VRE, or a Social Science gateway, that is cross

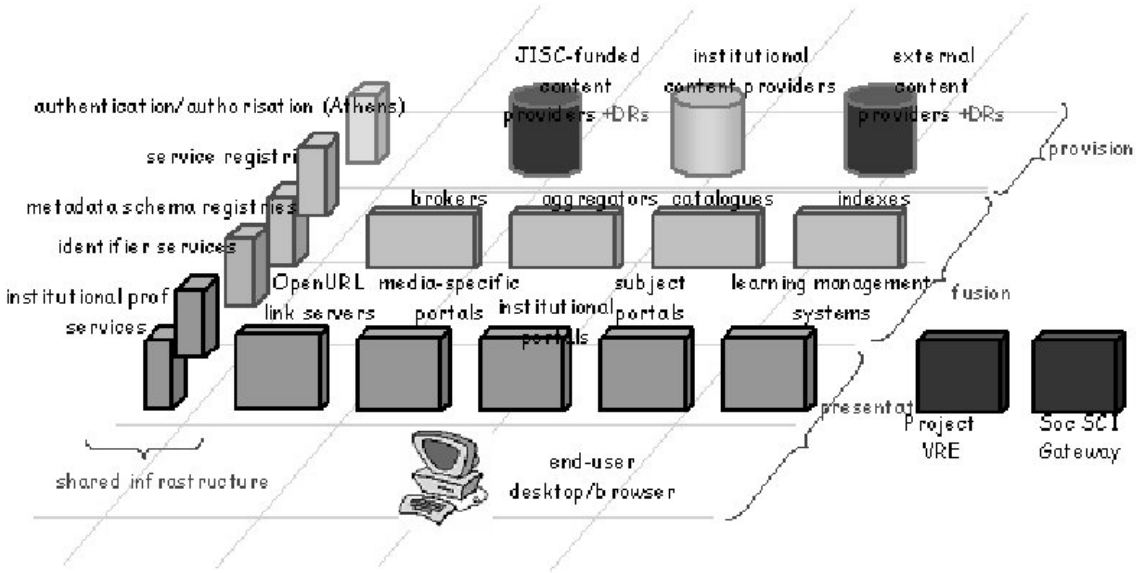

Figure 4: IE Architecture version 2

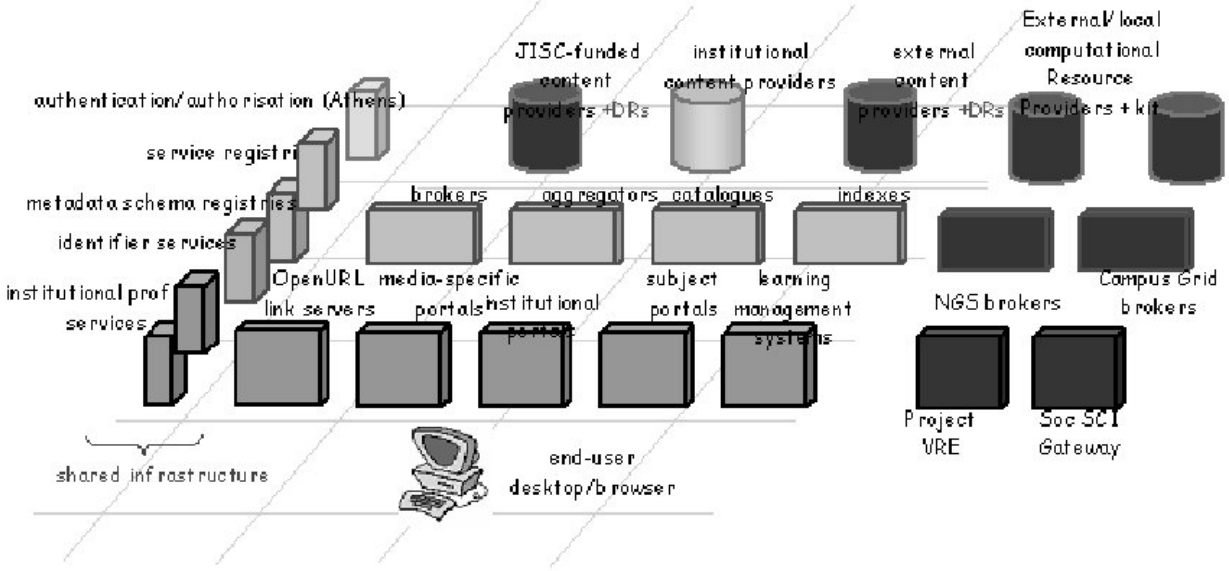

Figure 5: IE Architecture version 3

connected to the DRs. It could be just a browser, but that would lack functionality, so we need to add these to the front row. This part of the scenario highlights the need to link data sets to publications [2]. See Figure 4.

2. A this stage, the problem with Figure 4 is that it does not contain any computational facilities, so these have been added. Not sure what symbol to use for the actual kit, or whether this is not needed as its implied [3]. See Figure 5.

3. This stage requires the creation of new tools, which run on the computational facilities, so we need to be clear that its not enough to provide the physical infrastructure interconnections, there will be a bunch of new software tools and middleware. This will have implications for the choice of service components.

4. This stage is going to jointly use the DRs and the computational facilities.

5. This stage is going to use the journal literature (external content), but will require the use of new content harvesting and synthesising services/ tools.

6. This stage requires new tools as well that extend the functionality of those in stage 5 but add the data to the mix.

It therefore seems that it is not sufficient for us to simply add elements to the original IE Architecture diagram we also need to look at the "brick wall" service diagrams coming from e-Framework activities `http://www.e-framework.org` to see what elements we need for our scenarios and build the Research Reference Models.

We analyse the Social Science Research (SSR) scenario again with this in mind:

1. This uses the highlighted elements;

2. This uses the highlighted elements (4 new ones);

The other stages can use the same services accessible via a portal as part of the extensible VRE.

C What kinds of Portals will be met by Researchers?

The researcher is likely to meet Web browser-based portal technology in three situations: (1) the Institutional Portal provided as a gateway to the services and information of an institution or large facility and maintained by central IT staff; (2) a Project Portal with all the resources of a particular multi-institution research project – a Virtual Organisation – probably maintained by project staff part time; and (3) a Service (subject-specific) Portal provided for access to a specific service, e.g. a national data center, maintained by payed IT staff as part of the service.

[2] One JISC-fuded project investigating this is CLADDIER [8]. JISC is also funding links to data, e.g. on MIMAS and EDINA via the JCSR in the GEMS project.

[3] JISC is funding computational facilities for the support of research via JCSR, e.g. NGS, the National Grid Service [20]

The following definition is from Wikipedia `http://www.wikipedia.org`: *Web portals are sites on the World Wide Web that typically provide personalized capabilities to their visitors. They are designed to use distributed applications, different numbers and types of middleware, and hardware to provide services from a number of different sources. In addition, business portals are designed to share collaboration in workplaces. A further business-driven requirement of portals is that the content be able to work on multiple platforms such as personal computers, personal digital assistants (PDAs), and cell phones.*

Many of the portals started initially as either Internet directories (notably Yahoo!) and/ or search engines (Excite, Lycos, AltaVista, infoseek, and Hotbot among the old ones). The expansion of service provision occurred as a strategy to secure the user-base and lengthen the time a user stays on the portal. Services which require user registration such as free email, customization features, and chatrooms were considered to enhance repeat use of the portal. Game, chat, email, news, and other services also tend to make users stay longer, thereby increasing the advertisement revenue.

Different types of portal are defined to include: Regional Web Portal; Government Web Portal; Enterprise Web Portal.

C.1 Institutional or Facility Portals

Wikipedia goes on to say: *In the early 2000s, a major industry shift in Web portal focus has been the corporate intranet portal, or "enterprise Web". Where expecting millions of unaffiliated users to return to a public Web portal has been something of a mediocre financial success, using a private Web portal to unite the Web communications and thinking inside a large corporation has begun to be seen by many as both a labor-saving and a money-saving technology. Some analysts have predicted that corporate intranet Web portal spending will be one of the top five areas for growth in the Internet technologies sector during the first decade of the 21st century.* We might also refer to these as "Institutional Portal". They could be designed for or provide views for a variety of purposes: e-Learning, e-Research, Information Management, Administration, etc.

In this context Gartner defines "higher education" portals as *enterprise portals integrated with administrative, academic and other applications of interest to students, faculty and staff.* They place them high up on the "slope of enlightenment" in their 2005 HE hype cycle because, although budgetary constraints have slowed down adoption, they are emerging as key institutional interfaces for online resources and applications.

Many universities have started to develop portals, usually starting with a student portal and then moving onto other stakeholder groups, e.g. prospective students, staff, alumni. These can use portal software, e.g. Luminis, or can utilise the portal features of other enterprise software, e.g. Oracle or WebCT. Open source portals are in development, e.g. uPortal. Other organizations such as Research Councils are developing their own portals (e.g. ESRC Society Today, `http://www.esrcsocietytoday.ac.uk/`). STFC is investigating portals for access to large-scale experimental and computational facilities.

There were two institutional research portal projects being piloted under the JISC VRE programme. ELVI (Evaluation of a Large VRE Implementation) at Nottingham University `http://www.nottingham.ac.uk/research-systems`, and EVIE (Embedding a VRE in an Institutional Environment) at Leeds University `http://leeds.ac.uk/evie`. These sought to evaluate the embedding of research tools into

institutional portals.

Some features of enterprise portals are:

- Single point of contact – the portal becomes the delivery mechanism for all business information services (one stop shop);
- Collaboration – portal (institution) members can communicate synchronously (through chat, or messaging) or asynchronously through threaded discussion and e-mail digests (forums) and blogs;
- Content and document management – services that support the full life cycle of document creation and provides mechanisms for authoring, approval, version control, scheduled publishing, indexing and searching;
- Personalization – the ability for portal members to subscribe to specific types of content and services. Users can customize the look and feel of their environment;
- Integration – the connection of functions and data from multiple systems into new components/ portlets.

Most enterprise portals provide single sign-on capabilities to their users. This requires a user to authenticate only once. Access control lists manage the mapping between portal content and services over the portal user base. This is facilitated by a Corporate Data Repository within the institution.

C.2 Project Portals (Science Gateways)

Whilst an Enterprise Portal might be very good for e-Learning and Administration, as shown in the Lumenis demo, they provide an outward-facing representation of the processes and community within a single institution or organisation.

A Project/ Grid Portal used for e-Research will typically be used by people from many organisations. We will refer to this grouping of people and underlying resources as a "Virtual Organisation".

The logic underlying a Project Portal must facilitate sharing of data and resources within the Virtual Organisation which means across institutional administrative boundaries. Typically this requires Grid Middleware to comply with differing standards, policies and procedures.

C.3 Service and Subject-specific Portals

Service-based portals are now very common. Examples include Google, Amazon and e-Bay which are familiar to millions of people worldwide. They have many similarities to project portals, but are focussed on the end to end delivery of a specific service or set of services to its customers/ users.

There are many subject-specific portals, such as Arxiv `http://arxiv.org` (Cornell University), PubMed `http://www.pubmed.com` (NIH), or UKPMC: UK PubMed Central `http://www.wellcome.ac.uk/`

`doc_WTD015366.html` (Wellcome Trust). Many experienced researchers prefer subject-specific portals which contain deep-search and other facilities which they can use based on specilist vocabulary and subject knowledge.