JISC Information Environment Portal Activity: supporting the Needs of e-Research. – Interim Report –

Rob Allan
CCLRC e-Science Centre, Daresbury Laboratory,
Daresbury, Warrington WA4 4AD

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

Caroline Ingram
CSI Consultancy Ltd., 42 Coquet Terrace, Newcastle upon Tyne NE6 5LE

Contact e-Mail: r.j.allan@dl.ac.uk, r.crouchley@lancs.ac.uk, caroline@csiconsultancy.co.uk

May 18, 2009

Abstract

This is an interim report of the *JISC Information Environment Portal Activity: supporting the needs of e-Research*.

The aims and objectives of this study are:

- To scope the requirements of e-Research within the area of resource discovery with reference to "portal" type services and tools;
- To identify gaps and duplication within the current provision (with reference to JISC portal and other relevant activities) therefore to identify potential areas for new work and possibly synergies that could offer a more holistic approach than currently available;
- To highlight issues and challenges that will need to be addressed in terms of serving e-Research requirements and in terms of enhancing portal activities for the IE more generally;
- To make recommendations for portal related activities that could be taken forward by JISC.

In this report we principally document the background and procedure to be adopted in the rest of the survey alongside a desk-based analysis of the JISC Information Environment and related activities as appropriate to e-Research. The rationale for our selection of projects to review is explained.

Contents

1 Summary

In this study we aim to:

- Scope out what portal activities are underway within the e-Research community and the Information Environment more broadly (with special reference to resource discovery and JISC funded portal developments and services);
- Identify the broad requirements for resource discovery and portals within the e-Research community;
- Identify gaps and duplication and potential areas for streamlining/ co-ordination;
- Highlight issues and challenges that will need to be addressed in terms of serving e-Research requirements and in terms of enhancing portal activities within the IE more generally;
- Make recommendations for portal activities that could be taken forward by JISC.

These items will be addressed in the sections below.

We provide separate documents giving background information:

- *Scenarios, Use Cases and Reference Models* [1]
- *Comparison of Surveys* [2]
- *Web-based Library and Information Services* [3]
- *The Information Environment and e-Research Portals* [4]
- *Interim Report* [5]
- *A Vision for a Portal access to Global Information* [6]
- *Final Report* [8]

2 Introduction and Background

2.1 Rationale

We will review the needs of the JISC VRE programme and a selection of RC-funded e-Research (e-Science) projects as listed in the OST document [16]. This review will focus on how the Information Environment can meet the needs of e-Research through the provision of portal-ready services.

In the proposal to JISC, we had suggested a questionnaire and series of interviews with stakeholders plus one or more workshops to seek input. We quickly realised that much of this work had already been done, and that documentation existed from which we could quickly draw important conclusions.

We have therefore taken into account a large amount of existing material, as listed in the references to the *Comparison of Surveys* [2]. This has speeded up our work, by reducing the number of interviews needed, but also means that we have had to re-focus not to duplicate this effort. We have prepared a separate document summarising the previous surveys and picking out conclusions of relevance to the current work. This has lead to the definition of a few additional questions which cover areas of interest to us which were not previously addressed. These questions are being posed to a members of the e-Research community via an on-line "card-sort" based questionnaire `???` and were raised at the workshop in Lancaster 6-7/9/06. [add URL] The card sort tool was developed in the ICONEX project at University of Hull. `http://www.iconex.hull.ac.uk`.

1. please indicate which of the RDN Intute survey [1] top 10 priorities you would personally find useful;
2. how advantageous to your research would it be if you were able to access all the resources you might need (be that data, publications, databases, images, or any other online resource) through one interface?
3. We have extracted a list of services identified in the RIN survey [2], can you describe the online services you used and any particular reasons for using them?
4. which of thes following services would you find useful and be likely to use through a portal interface to support your research or researchers in the area you support?
5. we have extracted a list of services identified in the ESRC survey [3], which of the following services would you find useful and be likely to use through a portal interface to support your research or researchers in the area you support?
6. please indicate your research or research support subject area from the following broad category list?
7. do you agree to undertake a personal interview face to face or by telephone?

In addition to surveying previous responses, related consultancy documents and carrying out a number of interviews with key players we have worked through a series of use cases [1]. These lead to a conceptualisation of the components needed in the IE and how a researcher might use them integrated with other processes, such as administration, on-line research, learning and collaboration. Some sample use cases are illustrated below but a bigger set should be sought to validate our conclusions.

2.2 Domain Differences

In attempting to focus the rest of our activities on the needs of e-Research we first identify some domain differences.

The "domains" or "silos" widely thought to demark current JISC activities are currently: (1) e-Research as managed by the Committee for the Support of Research; (2) e-Learning as managed

[1] [give ref to RDN Intute survey] `http://www.rdn.ac.uk`
[2] [A series of interviews were carried out by RIN `http://www.rin.ac.uk`]
[3] ESRC RIESS Survey (2006) [?]

by the Committee for Learning and Teaching; and (3) the Information Environment as managed by the Committe for the Integrated Information Environment [4]. There are others, but they are less relevant for this study. In fact JISC are striving to develop programmes which bridge these domains for economies of scale, benefits in sharing of practice, tools and information. One of the main aspects of the present study is to help to contribute to the breaking down of silos and sharing of information and services.

We outline one possible view of the difference between these which we believe has a bearing on the way activites may cross domains. JISC are working hard to identifiy commonalities and enable these domains to be bridged and for software and resources to be re-used. This is partly undertaken in the work of the E-Framework for Education and Research. The driving forces of these domains can be summarised as follows:

e-Research

Many researchers are carrying out activities to generate "new knowledge". They are active across a wide range of disciplines and their procedures vary enormously from one to another encompassing 'observation, experimentation, computer simulation and analysis of historical materials or disipline-related corpora. These activities are characterised as being diverse, complex, multi-organisational if collaborating teams are involved, and dynamic since they evolve rapidly as new hopotheses and procedures are developed. The actors involved tend to be peers. Access to scholarly publications is important to researchers, but so is creation and archiving of new and access to existing scientific data.

e-Learning

This is characterised by being largely institution-based with established and relatively well-defined procedures including pedagogy, administration and assessment in various forms. The actors involved can be arranged hierarchically in peer groups, such as staff and students. Access to digital resources explaining and outlining methodologies and their application to example studies is a cornerstone of this area.

Digital Information

This is characterised by a body of people maintaining and making available information – we will focus here on digital information. Digital information has many forms including for instance text, sound and video material. It can include scholarly publications and scientific data, in either raw or seconday forms but packaged with appropriate meta-data explaining at least provenance, format and location. The focus of the Information Environment is to provide services for the publication, discovery, access and to a limited extend analysis of this material (the latter through "generic" service such as mining or markup). The Information Environment recognises that there is a need to deliver resources in variety of ways and therefore to interface with different presentation services/ systems such as those that support learning and research.

There are also differences between science domains, as discussed in [1]. We will refer to these as "subjects" to avoid confusion.

[4] Responsible for development of middleware and AAA services, core common services such as resource discovery and curation, core interoperability standards, IPR and activities like the e-Framework for Education and Research. There are overlaps with other committees and some share parts of this agenda.

2.3 High-level Questions

In this section we list some general questions and statements which have arisen during our study.

1. The IE Architecture needs to be extended to accommodate components relevant to other research processes. What are the additional components and how can this be done?
2. Researchers do not know what the IE is [5]. It is unlinkely that researchers (other than those funded directly by JISC programmes) will support or contribute to the IE or actively link its services into their own work. There is thus a culture of "re-inventing the wheel", rather than outcomes of all programmes, whether from JISC or the Research Councils being integrated for a common goal. What can be done to improve this?
3. Digital data curation is important, but probably not recognised as such by many researchers at present;
4. Researchers do not use deep search facilities sufficiently, but tend to use Google as their main tool. Google has largely replaced the use of citation indexes. Can Google be used to lead researchers to other tools?
5. The IE needs to embrace the needs of researchers to manage and share personal information. Is this part of its remit, and if so what technology should be used?
6. There has to be more linking between publications and raw data from scientific studies. This is only appropriate when that data can be shared. Projects investigating this should be accelerated and the best solutions implemented in the IE. Providers of data archiving services need to be involved;
7. Open Archive initiaties are growing in importance, such as e-Prints and e-Pubs. The IE needs to be seen to be playing a leading or at least a strong role in this. JISC are already investing some £20M in OA thought the Digital Repositories programme, and this needs to be made more visible and its outcomes embedded in research practice;
8. There are many sources of information which are outwith the JISC IE. The World Wide Web is a major source of information and knowldge. There are also many proprietary sources. Search facilities must embrace all these sources which may require bilateral agreements to be in place;
9. What user interfaces do researchers want/ need? Are portals suffient or should services be provided which can be linked into applications and desktop tools?
10. Tasks of accessing and publishing information are only part of research-related "admin" procedures which are growing in complexity. It must be possible to tie systems together using open standards;
11. Security, confidentiality and IPR are major concerns;
12. Metadata is important, for instance provenance;

[5] JISC claim that the IE is a wide range of resources, standards and protocols, services and also projects and programmes but it is not really a "thing" and is more about delivering resources (and curating them) in a more standard and seamless way that enables interoperability, diversity and rich use.

13. Electronic publishing leads to new business models, e.g. payment for individual download of journal articles rather than publish of shared whole issues. The impact on the researchers needs to be understood and potential negative impact addressed;

14. Persistent URLs, registries and resolvers are key elements of maintaining an information infrastructure in addition to the underlying services.

2.4 Discussion of e-Research and the IE

By e-Research the JISC means both e-Science and research within the humanities and arts that is undertaken within a digital environment. Researchers require input from both e-Learning and the Information Environment, mainly in terms of discovering and structuring background information to form and guide their research. Support for research is also important in bidding, reviewing and managing projects. There is also an important publication/ dissemination phase by which researchers are sometimes judged, as published output or citations are quantifiable metrics.

Increasingly there is recognition that the digital information, e-Learning and e-Research communities have the potential to work together and to use common solutions. The Roadmap for a UK Virtual Research Environment [**?**] has already identified resource discovery as an identified common service. The activity within the resource discovery arena has been disparate and there is benefit in investigating the potential of how portals and portlets, within the Information Environment area, could be enhanced to support the needs of e-Research; and how the resource discovery approaches, currently in use within the e-Research community, might benefit the Information Environment and JISC portal activity more generally.

Andy Powell, formerly at UKOLN, has provided a number of reviews of and opinions on activities of the JISC Information Environment [13, 14, 20]. We draw upon this input below.

The IE is currently focussed on the provision of resource discovery and related services for digital libraries and other catlogues, e.g. museums and image collections as well as digitised reports, theses and books. There is some support for survey-based scientific data (e.g. for social science), geospatial data and data from the arts and humanities (e.g. music, digitised paintings and manuscripts). This is provided via services such as MIMAS, EDINA, UKDA and AHDS. There is a growing awareness of the important of bio-medical and legal sources, but there is currently little overlap with the natural sciences such as supported by the Research Councils and subject-specific services (mostly journals and learned societies) other than ESRC and AHRC.

There is however a recognition just beginning that archival of publications and links curated scientific data is important. This is already done my some learned journals, such as those in the field of protein crystallography, and is being investigated by open archival projects such as CLADDIER [25] and StORe [26]. Differences between the handling of data and information are noted in [1].

A final type of information which is not yet being considered is "personal information". This could be a strong focus of the ongoing VRE programme, which has identified tools such as Wiki, BLOG, Forum as being useful to collaboring teams of scientists. The sharing of personal information is likely to grow through the use of peer-to-peer systems. Currently mostly used for popular music and video systems such as Shareaza, FilePipe, Ares, LimeWire, BitTorrent etc. could be adapted for scientific use, see `http://www.zeropaid.com/`. There needs to be some way of including such information in

information discovery services but differentiating it from peer-reviewed content. There is as far as we can see only on project currently funded to investigate this; SPIRE [27] in the Digital Repositories Programme which is using LionShare and aimed at sharing learning objects.

2.5 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.

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.

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 [17]. Other organizations such as Research Councils are developing their own portals (e.g. ESRC Society Today, `http://www.esrcsocietytoday.ac.uk/`). CCLRC is investigating portals for access to large-scale experimental and computational facilities [28].

There are currently 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) [10] at Leeds University `http://leeds.ac.uk/evie`. These are seeking 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.

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.

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 used based on specilist vocabulary and subject knowledge.

Summary of Google functionality

The differences between a Google search and an IE cross search are explained by Powell [19] paras 3.5-17. He also describes a number of initiatives to provide more open search access to repositories which are not published as open HTML documents.

Whilst Google is very heavily used in the research subject many other information services can/ could be provided with equivalent portal interfaces. Google is primarily a search engine, but its success is gained from the range of sources it can index and the personalised ans specialised facilities it provides. A summary of Google's capabilities might be useful at this point.

Google: caching, archiving and conversion of source to HTML

Google Search Engine Appliance: search engine for institutional intrantes. Can be purchased or a special instance hosted. Yahoo offer similar services. JISC services could provide similar tools for searching local repository resources.

Google Scholar: A beta service that *enables you to search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research.* `http://scholar.google.com`.

Google Maps: `http://maps.google.com`

Google Earth: a geographical information system combining satellite images, maps and Google search `http://earth.google.com`. It requires download of client software (in beta test).

Web Service: In addition to a Web portal interface several of the Google services are also provided via Web services. A simple customer key protects against dinial of service and can provide for limited personalisation. This is important in a research context where Web browsers are not the only interface. JISC services should be encouraged to provide similar Web Service interfaces in additional applications which require browser clients.

Another somewhat different general resource that is growing in importance is Wikipedia `http://www.wikipedia.org`. As its name implies, this is an on-line encyclopedia developed using Wiki technology.

It is an example of "folksonomy" in action, where anyone in the world can add text. Its accuracy is assured simply because there are a very large number of contributors and editors and errors are quickly removed. It has for instance interesting entries for "Digital Repository" and for "Portal", the former linking to the JISC DR Review of 2005.

3 Use Cases

An analysis of some sample scenarios and use cases is provided in a separate document [1].

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 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.

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.
```

3.1 What Functionality does an e-Researcher need?

We assume that this functionality will be delivered via a Web portal, perhaps through tools in a Science Gateway or Institutional Portal. We note 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 Pro-

gramme 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 was illustrated above.

In our consideration of the IE 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 [20] 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.

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 [24]:

- 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 [10]:

- 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 the research life cycle. We will consider this further in our "vision" document [6].

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].

3.2 What components need to be integrated in the IE and other infrastructure?

[more explanation, clarify]

Components come in the form of: (1) software service interfaces which can be combined into designs for actual applications, including portals; (2) fusion-layer services, including security; (3) infrastructure such as data servers and computers which host the services. In this dialogue we envisage the use of a Services Oriented Architecture (SOA) paradigm consistent with the JISC e-Framework for Education and Research.

In a separate document [1], we discuss the reference models and implied additional components in the IE Architecture originally proposed by Liz Lyon and Andy Powell [13].

4 Portals and Related Activities

We scope out what portal activities are under way within the e-Research community and the Information Environment more broadly (with special reference to resource discovery and JISC-funded portal developments and services). A survey of individual IE services and activities focussing on the presentation layer is presented in a separate document [4]. This includes information about relevant e-Research projects.

4.1 The JISC Information Environment

The IE has initially targeted the following areas: Images; Geospatial Data; Moving Pictures and Sound; e-Learning; Journals, e-Prints and Scholarly Communication.

In the IE and portal survey we have attempted to identify successful projects which have delivered re-usable software components or services. A good example of this is HEIRPORT which originally developed a portal for geographical searches on historical information in ColdFusion. Outcomes of this work are being used in the Common Information Environment (CIE) and CREE. HEIRNET have a production portal using the original technology, which was extended in HEIRPORT2. Client and server software was delivered and is available to download together with documentation. The CREE project converted the servlet interface to a portlet to incude with other search facilities such as JAFER in a uPortal framework and also make available to other projects using WSRP. In the Sakai VRE Demonstrator project we have indeed shown that we can use such services.

The JAFER project (`http://www.jafer.org`) has itself delivered a cross platform toolkit in Java which can be used to interact with Z39.50 servers. However none of the demonstratinos mentioned on the Web site seem to have targeted the more widely used research resources.

CREE integrated HEIRPORT and JAFER with other resources accessible via GetRef, GetCopy and Google. Some of the portlets are available to download (currently only JAFER, but Google is to appear soon).

Other projects such as SPP, which delivered cross-search facilities and a variety of portlets for the RDN, have similar potential to be integrated with other services.

Other projects which have made their services available as portlets include Go-geo! and Xgrain (the project which produced GetRef). Both of these provide software which should be transferrable to the research domain, e.g. accessing Grid-based resources. [re-write]

[CONNECT for funding opportunities]

We note the recent Invitation to Tender for a JISC Information Environment Testbed. This reference testbed, if successful, is intended as a vehicle to demonstrate the vision of the IE with working infrastructure components and also enable new projects and other stakeholders to access it to test extensions and interoperability of the tools they are delivering. It will be particularly important for the testbed to implement precise versions of the protocols and standards being used. [explain why this is referenced more clearly...]

[Could these portlets be used to deliver or search large data sets? Are they scalable? are there int eroperability issues? (i.e. a perceived R need is access to this type of info.). We assume that th e supporting documentation will draw this out but we will need a clear view of what portals and por tlets are in common use for researchers.]

4.2 e-Research Activities

Some of these have portals, as noted in a separate document [4]. Whilst most are project specific, and might integrate interfaces to subject repositories such as in the Integrative Biology VRE or the

Histrory of Political Discourse VRE, others are more generic demonstrators. These include the Sakai Portal Demonstrator, EVIE and ELVI. It is likely that some of these will deliver software that can be re-used. An additional project of interest is ShibGrid which is extending portal frameworks to be able to use the Shibboleth federation for authentication of academic users. This will be included in the National Grid Service portal for access to tools for managing applications and data on the NGS.

4.3 Portal Frameworks and Standards

A survey of portal frameworks and standards is presented in a separate document [12]. This is an update of our technical report from 2003 [11]. A very useful introduction to this was sent to use by Anurag Shankar of Indiana University. His work is part of the US TeraGrid activity. His introduction to portals is currently included on the ReDReSS Web site `http://redress.lancs.ac.uk` under "learning space/ portals and portlets". It will also be included in the resource Wiki of the eReSS project supporting the VRE programme.

5 Gap Analysis

In this section we present a preliminary gap analysis. We will identify gaps and duplication and potential areas for streamlining/ co-ordination. This will be extended with suggestions of how the gaps can be filled in the final report. Some of these gaps were infact noted from the survey of the StORe project and in the UK e-Science Gap analysis [15].

Common metadata model: useful for registries and discovery tools, especially in cross searching. Current subject-specific or *ad hoc* models have overlapping or exclusive terminology and semantic support (e.g. ontologies) is needed to enable inter-operation;

Cross search: genuinely integrated cross search facilities rather than just a set of links to other tools;

Common access standards: once discovered, information has to be accessed and consumed – cross searching and merging which we term "marshalling". There are data format and packaging issues;

Linking: flexible end-to-end links are required between data and information with a "citation" mechanism updating the links;

Personal Information Management: tools for archiving and sharing personal information are often requested. Some systems provide a shared "bookmarking" capability such as Connotea. Others offer peer-to-peer services for publishing information in ways less formal than Web based open access archives;

Non browser-based Clients: provide access to a wide range of "active" research resources, not just from a Web browser;

Mobile Research: access to information systems from mobile, context-aware devices, for instance decision support in extreme conditions or emergency situations;

Awareness: Most (?) researchers are not aware of the JISC activities and do not always ask specialised library staff for help with information searches.

Culture: projects must produce tools which work through established hierarchies and communication routes.

In addition to technical gaps, there are other gaps which have been noted in a number of surveys. An important one is on ICT training for end-users of information systems, see [2].

6 Recommendations

Here we highlight issues and challenges that will need to be addressed in terms of serving e-Research requirements and in terms of enhancing portal activities within the IE more generally. We make recommendations for portal activities that could be taken forward by JISC. These will be prioritised to a more finite set in the Final Report [8] based on additional feedback.

Based on his analysis in 2005, Powell [20] made the following recommendations. We are aware that many of these recommendations are already being taken forward. Comments from our knowledge of the e-Research landscape are given as footnotes.

1. The JISC community needs to work to ensure that the service oriented approaches being adopted by initiatives like ELF, the DLF framework, Sakai, VIEWS, etc. will use the same conceptual frameworks and terminology as far as possible [6].

2. Engagement with the JISC IE by the commercial sector and other players is extremely valuable to the community and we should take care not to lose this important buy-in to our shared activities as we move forward with a more service-oriented approach [7].

3. The community needs to increase its investment in automated approaches to metadata creation and automated approaches to indexing and data-mining full-text and multimedia resources. We also need to remember that there will probably always be scenarios for which manually created metadata will be the most appropriate solution.

4. The JISC community needs to maintain good links with the Semantic Web Best Practice and Development Group and with other key players, particularly in the areas of metadata schema registries and terminology services [8].

5. Development of services that support community-driven approaches to building terminologies, so-called "folksonomies", are worthy of consideration for JISC-funding. Evaluation of these kinds

[6] The e-Framework for Education and Research is a formal approach to defining and classifying such services. It is hoped that the definitions will be accepted to all domains and that underlying services can be re-used and integrated with domain-specific services. In this way IE services can be directly integrated with e-Research tools.

[7] There is much less commercial presence in the research domain because of its diversity.

[8] Some other issues noted by Andy Powell concern the automatic production of metadata and use of Semantic Web technologies. These have been the focus of several e-Science projects and there may be a useful synergy and transfer of technologies between e-Research and IE.

of approaches would also be useful to the community [9].

6. JISC should encourage the community to experiment with peer-to-peer (P2P) approaches (within single institutions, between a limited number of institutions and nationally) in order to gain some experience of their strengths and weaknesses [10].

7. The community needs to develop a typology of "repositories" (eprint archives, institutional repositories, learning object repositories, content management systems, etc.) in order to understand their differences and similarities and in particular how to enter into appropriate dialogue with the commercial sector about the supply of software to deliver them.

8. The JISC community needs to collaborate internationally on the modelling of "complex objects" and their packaging using standards such as METS, MPEG-21 DIDL and IMS C/P. Furthermore, the community needs to build an infrastructure that provides a coherent view across disparate repositories in order to prevent individual service providers having to replicate significant pieces of knowledge engineering [11].

9. The JISC community needs guidance about how best to expose the content in repositories to search engines like Google, whilst at the same time also investing in more structured disclosure approaches such as those based on metadata harvesting and cross searching.

10. The JISC community should work with selected content providers and end-users in order to undertake some appropriate research into the effectiveness of exposing full-text to Google and metadata to metasearch engines and the end-user benefits that such exposure brings.

11. The JISC community should work with the NISO Metasearch Initiative and appropriate e-Learning partners to evaluate the use of the A9 OpenSearch specification, SQI and other similar specifications as alternatives to Z39.50 and SRW/SRU.

12. JISC IE content providers need best-practice guidance for how to assign relatively persistent http URIs to their resources and on when it is sensible to buy into alternative identification systems such as the DOI.

13. JISC should work with the providers of union catalogue services to investigate their use as points of contact with Google, both as places where metadata records can be exposed and as places where knowledge of physical and electronic holdings information can be disclosed. However, this work should not be undertaken unilaterally within the UK.

14. The community needs to see more development undertaken in the area of automated indexing and data mining of full-text and other content types. JISC should work to ensure that there are appropriate links in place where institutions are deploying full-text indexing techniques, e.g. in the provision of a university's Web-site search engine, with other institutional activities such as the development of e-Print archives and/ or institutional repositories.

[9] There is debate about this point. Development of formal ontologies by established experts is likely to be more useful but is time consuming. Several e-Research projects are making progress in subject specific areas. Projects such as Wikipedia have however shown that a folksonmy approach is not impossible but would require the widest possible community input.

[10] Some, but only a few, e-Research projects are evaluating this approach. The Sakai VRE Demonstrator is one. We noted it in our gap analysis.

[11] Also noted in our gap analysis

15. In order to support seamless resource discovery approaches across the content of repositories and to support personalised views of this content, JISC should investigate the benefits of developing a national federated architecture for repositories, in tandem with similar national initiatives elsewhere as appropriate. This would include agreeing common solutions to a variety of technical challenges such as the assignment and resolution of identifiers, the use of complex object packaging standards and the provision of format conversion tools.

16. The community needs to refine its performance measures for purely machine-oriented services such as those found in the fusion layer.

17. The community needs to balance the focus on portals as Web-based services with a focus on the most effective mix of desktop and Web-based tools and services (both machine-oriented and human-oriented) that can be used to meet end-users' functional requirements.

18. The JISC community should contribute to the development of a global OpenURL resolver "routing" service in order to encourage and streamline the deployment of OpenURLs on a very wide global scale.

19. The community should ensure that appropriate authentication, authorisation, and trust mechanisms are in place to support the potentially complex relationships between end users, institutions, shared services, fusion layer services and content providers.

20. The community should attempt to reach agreements internationally about how to deploy distributed "service registries" – including agreements on metadata standards and transport protocols. We also need to agree on the operational policies for service registries and the ownership and IPR issues associated with the metadata records being exchanged.

21. The JISC community needs to undertake more work in the area of mapping metadata schema and related services, looking particularly at the issues of mapping between Semantic Web and non-Semantic Web schemas. The JISC community also needs to consider setting up a registry of "packaging profiles".

22. The JISC community should continue to contribute to international discussions about the use of identifiers and the services associated with them [12].

23. As a community we need to refine our understanding about the best ways that our ontologies can be created and maintained and the kinds of services that we require on those ontologies. We also need to ensure that best-practice guidelines are developed for assigning identifiers to terms in the vocabularies (e.g. URIs) and for marking-up the vocabularies in machine-readable forms [13].

24. The JISC community should work towards building a licence registry (or registries) to encourage a consistent approach to the deployment and use of "open access" licences [14].

25. The JISC community should encourage the development of automated metadata-creation tools and should deploy them as Web Services so that they can be embedded into presentation layer (and other) tools and services [15].

[12] This is also important for data curation.
[13] This is very important in the research domain.
[14] This is also very important for licensed applications on the Grid.
[15] A lot of work is already going on in this area.

26. In order to deliver name authority services, JISC should work with various parties, including the BL, to determine if an "authority list" of journal article author names exists or can be created, and if so to layer Web services in front of it. Alternatively, the JISC community could consider options for delivering a distributed "name authority" service through a network of institutional (LDAP) servers.

The above personal comments were those of someone on the inside facing out.

A useful prioritisation came from the RDN Intute survey [21]. This was in response to the question to researchers *How useful do you think each of the following features would be?*.

Rank	Service
1	A directory of bibliographic databases in each subject
2	A directory of journals, with indicators of subscription numbers and prestige
3	A database of sources of funding and scholarships
4	A directory of regular conferences, with indicators of size and prestige
=5	Searching and browsing filtered by the time period which the resource concerns
=5	Ability to filter out resources not directly applicable to research
7	Online events calendars - for conferences and other events in each subject area
8	National directory of researchers and research
9	Data that can be exported for use on other Web sites, or for other uses elsewhere
10	List of resources for graduate skills training – publishing, networking, thesis writing, etc.

Another list of features, in order of priority, was presented in Section 4 of the User Requirements survey from the EVIE VRE Project [10]:

1. I want to see all of the resources and databases that are available, and select which are included in each search
2. The University's list of publications should include links to the full documents
3. Have a service that is easy to use, with little need for help
4. I want to monitor financial expenditure on my grants
5. An email alert containing information that I specify when it becomes available
6. A single Google-style search box to search across many resources
7. A What's New page when you logon to the Research Portal
8. Access software and information while off-campus
9. A single place where I can submit research results for long-term storage and easy access
10. Sharing and handling large files
11. Group diary and meeting organiser
12. Ability to search, view, and download proposals that have been previously submitted
13. Mechanism to find researchers with specific expertise within the university

14. An advanced search form enabling search on specific data fields
15. A structured bid template with electronic routing/ distribution and sign-off by University administration
16. A quick and simple way of submitting my research results to be archived
17. Access to documents and hyperlinks should be maintained even if staff transfer to new jobs or organisations
18. Automated alerts about new funding opportunities based on my preferences
19. Files to have version control so that they can be safely changed under multiple authorship
20. Control over who can view or modify shared files

Further details and options are given in the report. The report also has an appendix which lists comments received giving barriers to some research tasks. These include perceived lack of: funding, time, support, ease of use, multiple tools/ sources, no focus, no interest...

Prioritisations of this nature might guide future VRE developments to make use of IE and other services.

Our additional conclusions, from the opposite perspective, are:

1. Research covers a very large number of areas with diverse sources of information and data which must be accessible in various ways to suit researchers different modes of working in different disciplines.
2. IE therefore needs to broaden its content coverage to the natural sciences including raw and secondary data holdings. If this is not done, the uptake of the IE within the natural sciences will be largely limited to literature searches, many of which are currently done via Google or institutional library services, professional bodies such as IOP or a specialised scientific journal service such as Elsevier on-line.
3. Web browsers are only one client for research services and others, such as Web Services which can be linked into "heritage" applications, GUIs etc. iPod, Firefox, Matlab or other plugins are likely to be of increasing future importance. All information services (resources/ repositories) need a machine-to-machine interface to facilitate integration.
4. Existing IE services need to be exposed as "portlets" to be included in institutional, project or service (subject) portal frameworks alongside other functionality. This can be done via their M2M interface.
5. Once integrated into a single framework using M2M interfaces the cross-search tools must be created to pull information from the various services. This implies a need for meta-data standards, and potentially registries.
6. The portal frameworks in use by institutions for adminstration and e-Learning should be a focus for development. The open source version of uPortal would benefit from investment to make it more user friendly for e-Research. It is currently being used for the NGS Portal.

7. The IE is complementary to Google. It is unlikely that Google will be ousted as arguably the most popular generic Web search tool, but IE cross-search technologies could provide a much closer link between a research project portal and underlying subject-specific material. This is noted by Powell [19]. There is a clearly-articulated need for both generic keyword search and "deep" search capabilities.

8. The IE needs to address making available information about material worldwide, not just in the UK. This is because of the increasingly international nature of research collaborations. This may be facilitated by the adoption of standards and services as being defined in the e-Framework for Education and Research. This inludes access to information from publishers of learned journals.

9. Equally, IE has to ensure that publications from UK researchers are discoverable world-wide, this will improve the profile of UK research and citation ratings relevant to the RAE process and national prestige.

10. There needs to be more emphasis on enabling individual researchers to publish information at various stages, e.g. abstracts, pre-prints or full papers. Projects such as e-Prints and e-Pubs are beginning to facilitate this. IE should be a vehicle for integrating these activities. So called "private" information is also being shared in small and increasing circles of collaborating researchers using repositories such as Fedora or DSpace and peer-to-peer technologies. IE should seek ways to link to this kind of information with appropriate quality assurances and IPR considerations.

11. We are uncertain of the relevance of emerging technologies such as podcasting, which are now recognised in the educational domain. However we believe that IE should carefully watch this area as a means of information dissemination.

12. We believe that personlisation is important in portal frameworks, but have yet to address how this is to be accounted for in this study.

7 Acknowledgments

JISC for funding.

We thank people we spoke to: Ann Borda, Maia Dimitrova, Matthew Dovey, Bill Olivier, Andy Powell, Ian Dolphin, Chris Awre, Robert Sherratt, Debbie Franks, Catherine Jones, Penny Windebank, Roddy MacLeod, Graham Pryor, William Kilbride, Colin Nave, Dave Meredith, Mike Gleaves, Ruth Kirkham, Simon Hodson, Paul Beckett, David Gavaghan, Matthew Mascord, Sharon Lloyd, Graham Klyne, Stephen Lyon, David Zeitlyn, Michael Fischer, Janet Bagg, Simon Coles, Derek Sergeant

References

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

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

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

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

[5] R.J. Allan, R. Crouchley and C. Ingram *Interim Report* (CCLRC, June 2006)

[6] R.J. Allan, R. Crouchley and C. Ingram *A Vision for Portal access to Global Information* (CCLRC, June 2006)

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

[8] R.J. Allan, R. Crouchley and C. Ingram *Final Report* (CCLRC, June 2006)

[9] C.S. Ingram *IE Inventory with Images* (CSI Consultancy, 2005)

[10] 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)

[11] R.J. Allan, C. Awre, M. Baker and A. Fish *Portals and Portlets 2003* NeSC Technical Report UKeS-2004-06 `http://www.nesc.ac.uk/technical_papers/UKeS-2004-06.pdf`

[12] R.J. Allan, C. Awre, M. Baker and A. Shankar *Portals and Portlets 2006* in preparation

[13] Liz Lyon and Andy Powell *A Distributed National Electronic Resource Technical Architecture: Scoping the Information Environment* (UKOLN, University of Bath). See also `http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/`

[14] A. Powell *JISC Information Environment (IE) Inventory with Diagrams* (UKOLN, 2005)

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

JISC Virtual Research Environments Programme `http://www.jisc.ac.uk/index.cfm?name=programme_vre`

[15] G. Fox and D.W. Walker *e-Science Gap Analysis* (UK e-Science Programme, June 2003) `http://www.nesc.ac.uk/technical_papers/UKeS-2003-01/`

[16] N. Wiseman etc al. *Report of the Working Group on Virtual Research Communities* for the OST e-Infrastructure Steering Group (April 2006)

[17] uPortal from JA-SIG `http://www.uportal.org/`

Sungard SCT `http://www.sct.com`

[18] X. Yang, A. Akram and R.J. Allan *Developing Portal/ portlets using Enterprise Java Beans for Grid Users* Concurrency, Practice and Experience. Special issue on Portals. (2006, submitted)

[19] Andy Powell *The JISC Information Environment and Google* Discussion paper for the JISC (UKOLN, November 2004)

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

[21] RDN Intute Researchers Survey (2006)

[22] `http://www.rin.ac.uk/?q=search-discovery-access-menu` `http://www.rin.ac.uk/?q=user-behaviour-perceptions-and-needs`

[23] *Science Gateways* workshop at NeSC 19/5/06.

[24] OSSWatch Wiki

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

[26] *StORe: the Source to Output Repositories* See `http://www.jiscstore.com/WikiHome`

[27] *SPIRE: Secure Personal Institutional and Inter-Institutional Repository Environment* `http://spire.conted.ox.ac.uk/cgi-bin/trac.cgi`

[28] A. Akram, D. Choham, D. Meredith and R.J. Allan *CCLRC Portal Infrastructure to support Research Facilities* GGF14 (Chicago, 26-30/7/2005) Science Gateway Workshop