Arts & Humanities
Research Council

RePAH: A User Requirements Analysis for Portals in the Arts and Humanities

search
ortals
rts
umanities

Final Report

Stephen Brown, Robb Ross, David Gerrard, De Montfort University
Mark Greengrass, Jared Bryson, Sheffield University

September 2006

dmu.ac.uk
DE MONTFORT
UNIVERSITY
LEICESTER

The
University
Of
Sheffield.

Contents

1. Acknowledgements

We are very grateful to the hundreds of research practitioners in the Arts and Humanities who devoted time and care to answer our online questionnaires, attend our focus groups, respond to our telephone interviews, and answer our requests for additional information. This report was commissioned by the AHRC ICT in Arts and Humanities Programme, directed by Professor David Robey. Professor Robey and the staff of the programme have given us their assistance at each stage in undertaking the work of the Project.

The RePAH Project has necessarily involved examining the evidence relating to user-needs and user-behaviour collected or collated by the Arts and Humanities Data Service (directed by Sheila Anderson) and by the relevant branches of the Resource Discovery Network (now Intute), formerly Humbul (directed by Michael Fraser) and Artifact (directed by Colin Harris). We therefore owe a special debt of gratitude to them and to their staff, including the directors of the constituent branches of the AHDS (Julian Richards; Matthew Woollard; Mike Pringle; Martin Wynne).

We have also benefited from sharing data with two other projects in the ICT in Arts and Humanities Programme. Dr Lesly Huxley and the team of 'Gathering Evidence: Current ICT Use and Future Needs for Arts and Humanities Research' [University of Bristol] has been involved in a complementary exercise to assess user-needs in a different context. We have shared relevant information with them. With Dr Claire Warwick and the team of the LAIRAH Project ('Log Analysis of Internet Resources in the Arts and Humanities') we collaborated in our deep-log analysis, using the CIBER [Centre for Information Behaviour and Evaluation of Research] Team based at University College, London, directed by Professor David Nicholas. At the beginning of the project, Dr. Tom Leng kindly began work on collecting available published data for us. We are particularly grateful to Dr Paul Huntington, Senior Research Fellow in Data Mining and Web Metrics of University College, London for assisting us with our deep-log analysis.

The Steering Group for the RePAH Project met four times in the course of the twelve months of the Project. The following were its members, and they gave valuable time and advice at various stages of the Project, for which we are especially grateful:

Dr Matthew Woollard – *AHDS History, University of Essex*
Dr Michael Fraser – *Humbul, University of Oxford*
Jayne Burgess – *Artifact, Manchester Metropolitan University*
Dr Claire Warwick – *LAIRAH, University College, London*
Alastair Dunning – *AHDS, King's College, London*
Alun Edwards – *Humbul, University of Oxford*

2. Executive Summary

2.1 The Work of the Project

In July 2005, the RePAH Project was commissioned to carry out a survey of user-needs for information portals in the Arts and Humanities by the AHRC ICT in Arts and Humanities Programme. It began its work in September, conducted its first round of focus groups in December, also launching its online questionnaire that same month. By May 2006, the Project had analysed the 128 questionnaire responses, completed the report on the first round of focus groups and conducted a Delphi exercise among selected respondents. At the same time, deep-log analysis was conducted on the extant web-log information, mainly based on information from the calendar year 2005, furnished by the AHDS and two constituent elements of the RDN, Humbul and Artifact. This information formed the platform for a second set of focus groups, focusing on a 'demonstrator' of possible information portal developments. The responses to this second set of focus groups enabled the project to provide the fine-grained analysis of user-need which constitutes the basis of its recommendations. This report was compiled in August 2006 and submitted in September 2006.

2.2 Project Aims and Objectives

This was an **information-gathering** project. Our brief was to discover **user-behaviour** and **user-needs** of researchers in the Arts and Humanities in respect of portals. We set out to discover four kinds of information:

1. Information about users' information discovery strategies and internet usage.
2. Information about users' awareness and attitudes with respect to currently available online services and tools, including such gateways and portals as currently exist.
3. Information about patterns of recent user-activity in relation to the RDN subject hubs and AHDS.
4. Information about users' responses to what future portal developments can deliver.

2.3 Conclusions

Our initial analysis of the Arts and Humanities Research Community's research behaviour was substantially confirmed. This is a community which is **non-homogeneous**, **institutionally diverse** and **variegated** in its research patterns. We estimate it as around **50-60,000 active practitioners**, composed of the 'stakeholders' identified in our report – Postgraduate [PG], Postdoctoral [PD], Research Assistants [RA], Faculty and Independent Researchers [RI]. Our 'road-map' of their research

activities indicated a **core** and **penumbra** of activities, which are both **individual** and **group-based** [**A3**]. Not all these activities are currently universally served by the current information resource-discovery channels.

2.3.1 Users' information discovery strategies and internet usage

We emphasise the following features underlying Arts and Humanities research behaviour as regards their digital resource-discovery and information needs:

- o Digital resources are now **ubiquitous** for Arts and Humanities research. They are used **extensively**. Researchers believe that they have **fundamentally altered** the way in which they undertake research – i.e. the formulation of their research questions as well as gathering materials for answering those questions. At almost every stage of the research process, digital resources have changed the way in which Arts and Humanities research is now conducted. It has not yet, however, affected the way in which Arts and Humanities **publication** is conceived (although many journal papers end up on the Web). It has not fed through to the habits and procedures for **personal data archiving** nor has it had a substantial impact on the **means of scholarly communication** in the Arts and Humanities.
- o Our researchers emphasised that their agendas were **flexible**, **open-ended**, established on a predominantly **personal** basis, and **multiple**. They did not regard themselves as working in hermetically-sealed specialist areas. Rather, they saw themselves as researching overlapping domains, in which there were a series of core issues which could be tackled from a variety of differing angles.
- o Our researchers are **practical-minded** and **instrumental** in their resource-discovery strategies. The patterns were quite **discipline-specific**. Their needs are **extensive** and **broad-ranging**, reflecting their agendas. They expect their research methodology to involve a high degree of proficiency in resource-discovery. Our users are not promiscuous, but they have formed views on the perceived cost-benefits of using particular resource-discovery tools and strategies. These views are necessarily based on a sometimes less than perfect appreciation of the possibilities and range of a particular tool or digital library and of the possibilities of ICT generally. Both the questionnaires and focus groups highlighted a demographic within the arts and humanities community. There is a clear minority of scholars who are fluent in the use of digital applications and a sizable majority who find little need and/or time to use such tools.
- o All machine-accessible resource discovery depends upon implied taxonomies of knowledge. **Pre-structured knowledge** is not always greatly appreciated, however, by Arts and Humanities scholars. Their need for **assurance** about the authority and trustworthiness of a particular digital resource is in tension with the assumption that the **individual researcher has specialist skills** with which to assess its authority, by a **suspicion about who is undertaking the authentication**, and by an **awareness of the complexity that such a process entails**. They want to know about who has undertaken the authentication, and how often it is updated. They learn about the reliability of digital resources mostly from other practitioners, using established and informal lateral means of

communication within specialist fields. Arts and Humanities researchers are as likely to want to develop their own resource discovery trajectories as to follow those dictated by others.

o Categorised information is often not ideal when one is interested, as Arts and Humanities scholars often are, in the particular, or the anomalous. Resource discovery can provide pointers in the right direction, but Arts and Humanities researchers readily accept that **individual resource discovery** is fundamental to their research. The reiterative processes that this involves are a key constituent in the pursuit of, and definition of, their research agendas. Since Arts and Humanities research is still mainly defined at an individual level, information resource tools have therefore to be based upon these individual needs.

o There seems to be a significant relationship between the relative **lack of 'mutual dependence'** among Arts and Humanities researchers, their **'task uncertainty'** and the ways in which digital resources are **produced and utilised**. The arrangements for collaborative research and for disseminating research results are **personalised, localised and decentralised**. Informal communication depends on individual groups and specific social networks. Digital resources, where they exist, tend to be field-based and similarly localised. Likewise, there is a corresponding reliance on commercially produced generalist digital resources. We could produce no reliable estimate of what proportion of resources were in proprietary (i.e. commercially-provided, subscription-based or purchased information) as opposed to public-domain (i.e. free to access, generally publicly-funded information) information. Our users were often not aware of the contractual basis on which the information was provided to them. Nor could we estimate how frequently, and for how long, they consulted these resources – the patterns were too varied.

o There is a perception among arts and humanities scholars that within their fields there is little or no collaboration. The reality is substantially different, because while **strong collaborative cultures may not exist**, however, weak ones do and take the form of citations of colleagues' works, routine email correspondence, interaction through conferences and professional society meetings.

o Arts and Humanities **'e-infrastructure'**, apart from the AHDS and RDN subject-portals, tends therefore to be determined at the level of the **employing institution** rather than the field or discipline, or higher.

o We have to take into account a **'counter-culture'** (which we encountered most noticeably in our investigation of some of the disciplines in the Arts, where information resource needs and research agendas are often articulated in terms of 'diametrical difference' to prevailing trends).

o Arts and Humanities researchers want access to information, irrespective of the medium in which it is available. They are used to working in fields where there is a very **mixed economy of resources**, electronic and physical. Journal articles are important, but so are printed books. E-prints (pre- and post-) are markedly less significant than in the physical sciences and engineering. Electronic bibliographical information is therefore of critical importance to Arts and Humanities researchers.

o There is less emphasis on communicating work-in-progress and more emphasis on **formal ways of disseminating information**. There is consequentially less emphasis on lead-times for accessing research-sensitive information and results.

2.3.2 Information about users' awareness and attitudes with respect to currently available online services and tools, including such gateways and portals as current exist.

In general, we encountered a **high and growing level of expectation** as to the availability of materials in digital form. These expectations have been fed by the **exponential growth in the content of Arts and Humanities digital libraries** by the wide variety of different content-creators and contractors.

Generally users were largely **unaware of the possibilities** for data analysis and multimedia data presentation that digitisation offers and were equally **unaware of the extent to which their use of digital resources is tracked and analysed** by content and service providers and employers.

The **internet search engine** emerges from this study as an immensely useful digital resource-discovery tool. Users deployed a variety of proprietary search-engines. Their simplicity and speed appealed to our users, for whom a key determinant in their cost-benefit analysis of resource-discovery tools was whether it saved, rather than cost them time. At the same time, our users were also aware of the limitations of their internet search-engine of choice. Our users told us of their **frustration at its lack of sophistication**. They were **suspicious of its ranking of hits** returned. They were **overwhelmed by the information redundancy** which often accompanies its results. They were, above all, concerned about the fact that search-engines do not search a great deal of digital content that is relevant to their needs; and, equally, they are frustrated by the lack of interoperability between different libraries of digital content.

The issue of **'access'** runs throughout our report. Access to online journals was most often raised; but it frequently occurred also in respect of proprietary digital content of various kinds, specific to particular disciplines. The issue was sometimes presented in terms of a trade-off in resource terms, with our users wanting to see the investment of scarce resources in widening the local access to digital content through licence and content purchase rather than increased investment in resource discovery. At the same time, our research practitioners were aware that 'access' was only fully beneficial when it was linked to enhanced resource discovery, and, in particular, interoperability.

Interoperability was another major theme running through our enquiries. It tended to affect some disciplines more than others. As digital content becomes richer and more diverse, so the independent platforms on which it is consulted multiply. As interoperability becomes more important, so the potential for a next-generation resource-discovery portal grows. While the AHDS and Intute allow their resources to be harvested by other services, **they do not themselves comprehensively harvest available metadata**. For the AHDS this is due to their remit of collecting 'from' not 'for' the

research community, while Intute-Arts and Humanities has indicated a general lack of useful metadata available. Intute offers Really Simple Syndication (RSS) news feeds that aggregate news and new collections. This is a form of service that is already appreciated by individual users. This would appear to be a more advantageous route for making data available to commercial harvesters than that provided by the Open Archives Initiative (OAI) metadata-harvesting. The latter has currently received only limited take-up within institutions and none to our knowledge by individuals.

2.3.3 Information about patterns of recent user-activity in relation to the RDN subject hubs and AHDS.

From the wide-range of resource-discovery services and tools used by Arts and Humanities scholars, we investigated user familiarity with and use of these two services in particular. The key feature of the **RDN subject-portals** is their **resource descriptions**. Although our users were clear about the potential importance of **authenticating** digital resources, they were not so sure about the **significance** of the resource descriptions provided by the RDN portals. In particular, they had no sense as to how often they were **up-dated**, the **status** of who had written them, and what **range** of resources they covered. Those that had used the subject portals, took the view that they tended to be useful at the beginning of a research enquiry, but to become rapidly less relevant the more one advanced into a subject. Those that had not used the RDN subject-portals but knew of their existence had evidently formed a view about whether they were **likely** to find anything of relevance to them within it. We conclude from our evidence that the RDN portals are insignificant for most research purposes for the Arts and Humanities practitioner.

AHDS has a similarly low profile among the majority of arts and humanities researchers, although the evidence from AHDS web-logs may well be deceptive. Overall they may under-record some aspects of its usage despite some inflation of usage figures resulting from the inclusion of internal traffic between different servers within the AHDS network as a whole, including network administration calls. Although the number of **resources downloaded** seems to be increasing, none of the participants in our focus groups or questionnaire admitted to having downloaded such collections. Where the AHDS **harvested** data, generally in collaboration with outside partners (as in e.g. Heirport the Historical Environment Information Resources Portal), it plays a significant, perhaps pivotal, role in Arts and Humanities research.

Neither service has a published strategy for consulting users and discovering their needs, although there are examples of good practice in some parts of the AHDS. There are some good collaborative links with other information service-providers in place, but these need to be strengthened. The two services are not currently interacting very well. The RDN subject-portal does not harvest the metadata on AHDS resources comprehensively. While references to each other can be found on their respective sites, neither service promotes the other particularly actively, explains their relationship/differences or provides a quick and easy link to the other.

2.3.4 Information about users' responses to what future portal developments can deliver

Users generally found the current resource-discovery arrangements and services adequate, but were confused about their roles. The evidence is that researchers are more concerned with access to content than functionality.

At the same time, they recognize that the current situation with regard to functionality is not sustainable in the longer term. The importance of interoperability in users' minds was a measure of that realization. The exponential growth in data volume, combined with increasingly complex multilayered information, will make it more necessary to use resources in a complementary way, and simultaneously harder to do so.

Our users responded positively to the possibilities of a **personally-managed** research environment. There were specific, realizable functionalities that they identified as being of direct use to them in carrying forward their research agendas: **workflow management tools** and **resource discovery tools**. Researchers wanted greater personal control over digital resources. They readily perceived the advantages of tools which enabled them to integrate searching the web with searching their own hard-drive. They saw benefits to more developed **bookmarking** features, **personal editing** features, and an **automated copyright management system**. They wanted to be able to **filter** the quality of hit returns, **search distributed databases.** They responded positively to a **web-based news feed feature**, and liked the idea of **RSS feeds** that by-passed personal email accounts.

They were less excited about tools to enable communication and collaboration. The picture that emerged is of researchers who find asynchronous and largely mono-media communication channels such as email, web pages and telephone quite satisfactory. Real-time communications media such as instant relay chat and Grid videoconferencing with integrated computer applications sharing were less appealing. However most respondents declared themselves **happy to collaborate at the basic level of sharing the sources they used.**

Many of the features presented in the demonstrator imply a more sophisticated portal tool than the current gateways provide, and that requires a development in the ICT skills-base of the user-community which it is clearly reluctant to make. The investments made in the ICT skills-base through the Methods Network, ICTguides and training/awareness programmes organised by the AHDS cannot be expected to uplift the skills-base of researchers who do not currently see the need to do so. Whilst this skills-base is likely to improve over time, the potential functionality of portal tools will probably always outstrip it.

2.4 Ways Forward

We see a number of ways forward.

1. An awareness of the **distinctive research culture** with its fears and predilections must be taken into account.

2. The Arts and Humanities research community is not very assertive. Its digital resource-discovery needs have not been very well-voiced. As digital data expands exponentially in our field, and becomes increasingly complex and multi-layered, it is going to become harder to find, and use what we need. The arts and humanities need **strong pan-institutional organisations** that can champion the disciplines nationally and internationally. This is a role that AHDS is beginning to play in relation to standards (Brown et al 2006) but it applies also to information resource-discovery needs, including issues of access to content. The AHDS has a singular focus on arts and humanities. Intute offers a more integrated service of resource discovery within which Intute: Arts and Humanities has been established to function as a distinct service for the arts and humanities. The case for a single and coherent resource discovery service for arts and humanities is from the point of view of the user, clear.

3. The increasing provision of **metadata-harvesting** among the information service-providers is an immediate and short-term objective, dominating the agenda of resource-discovery over the next five years. Users are coming to expect much better **linkage** between **online bibliographical resources**, and the **online content** itself. They also want to search across **distributed digital data**. This objective implies:
    o **common metadata standards** [substantially in place]
    o agreed **authentication systems** [emerging, but more work needed]
    o much greater degree of **collaboration** among a wider group of information service-providers than is currently in place (research libraries: archives: museums: government/commercial information-providers, etc) [not in place]
    It is beyond our remit to recommend where such collaboration should come from. But we are convinced that the AHDS has a more important role to play in participating in, and facilitating, such collaborations than it has played in the past.

4. In the medium and longer term (in a five-ten year perspective), it is likely that the **semantic web**, especially when combined with harvesting agents, will provide the easy-to-use tools that many researchers need, at least to some degree. However, for some areas of the Arts and Humanities where "knowledge" is more the result of heuristics and associative thinking, it may be that a more folksonomic approach as exemplified by **Web 2.0** services such as Flickr and steve.museum will be more effective. We are therefore more persuaded in the shorter-term of the possibilities of **Web 2.0** offering a way forward in the form of community-contributed and mediated content. Users do not seem averse to contributing in

that way, but the nature of 'mediation' should be recognized. We can see the possibility of the RDN subject-portals evolving towards **a different mediation role**, with resource-discovery content coming instead from the community itself. In the longer term, there may be a possibility for combining the semantic-web and Web 2.0 approaches, especially if and where discipline-based ontologies emerge as commonly accepted.

5. We can begin to discern the determining characteristics of the resulting information environment as it emerges over the coming decade. It will be:

   o   inclusive
   o   aggregative
   o   personalisable
   o   locally managed
   o   quality-assured
   o   easy to use
   o   community-based
   o   internationally developed

At various points in this report we have referred to this as a **'managed research environment'**. The use of the term "environment" rather than "portal" is significant here because it does not necessarily entail a single provider. It could comprise a selection of Web portal services, or "portlets", that users draw down to their desk top and configure personally or it may take the form of a pre-configured set embedded within a trusted supplier such as an institutional or professional society web site. Moving towards such an environment should be regarded as a medium-term objective (i.e. three to five years). The current portal providers in the Arts and Humanities do not look like this. But, of course, there are already individual services in the public domain that have some or all of these features and there are recent precedents for the kind of environment we have described. For example, the JISC/LTSN Learning and Teaching Portal Project resulted in a set of web portal services that are embedded in the HE Academy website as a suite of 'Finder' services that could be adopted by other organisations (http://www.heacademy.ac.uk/48.htm).

We know that Arts and Humanities researchers are prepared to seek out and employ unusual, and 'unauthorised' sources for their information. We also know that they are willing to share useful sources they have discovered themselves. It seems likely that, if researchers come to recognize the existence and utility of such tools and services as these, they will employ them in greater numbers, further undermining the viability of established and 'authorised' services.

6. In the development of such a 'managed research environment' in the Arts and Humanities, there is also scope for collaboration with information system developers, including commercial and international providers. We do not exclude the possibility of UK collaboration in this area with developments currently under

Beta-test in 'Google Scholar' to share the costs and manage the delivery. Many of these tools will need to conform to the international standards that are encouraging British developments to be compatible with a much larger range of applications.

We therefore recommend a scoping study to ascertain the feasibility of such collaboration and the costs of developing a research-directed community-driven subject portal that offers:

- o **Workflow Management tools** that give the researcher greater personal control over digital project resources, especially more evolved **bookmarking features** and some form of automated **copyright management system** to facilitate the growing concern with usage permission and intellectual property rights was also highly valued.
- o **Resource Discovery tools** that provide greater control over web-based resources including the ability to **filter** the quality of hit returns, **search** multiple databases
- o **News feed features** that by-pass personal email accounts, but notify users of conferences, funding, jobs and new research publications.
- o **Collaborative research tools** for social bookmarking, uploading and sharing resources, annotating digital resources, shared document editing, attaching metadata to personally-created digital resources, and contributing to the authentication of digital content.

7. We recommend in the **short term** (one-two years) a much greater collaboration through data-harvesting of the current AHDS and former RDN subject-portals in resource discovery provision and though cross promotion of each others' services.

8. In the **medium term** (three-five years) we recommend that the AHDS and Intute develop a more Web 2.0 compatible profile to enable greater community involvement in resource recommendation, evaluation, creation, selection, sharing and annotation. We also recommend that funding bodies such as JISC and AHRC positively encourage and facilitate the development of interoperable portlets that can be used to embed portal type functionality in institutional and community web sites. An example of this may already be seen in the use of RSS news feeds offered by both services in order to announce news and collections.

9. In the **medium to long term** (five-ten years) we recommend that the AHDS and Intute: Arts and Humanities consider integrating their databases and user interfaces to provide the nucleus of a new, seamless, more comprehensive service in this particular area, one that combines and integrates the core functions of data-archiving, and digital resource harvesting/indexing.

3. Introduction

3.1 Background

How does the arts and humanities research community find and exploit the internet resources it needs? The question has no simple answer in terms of service provider. It is currently served by **complementary services**, each offering to act in some measure as resource discovery agents:

- o RDN subject 'gateways'. Their mission statement of 1999 was to construct a 'collaborative network which enriches learning, research and cultural engagement by providing a new level of access to high quality Internet resources'. The Arts and Humanities 'gateways' (Humbul and Artifact) – were merged into a single entity (Intute) in the course of our investigations.
- o The AHDS. The AHDS mission statement includes as one of its three planks: 'providing rich, deep access to the intellectual content of arts and humanities digital resources created by or for Higher Education.

These services offer **different resource discovery possibilities** to the user. Humbul/Artifact (now Intute) furnish collection-level descriptions about online resources and various 'value-added services' including online tutorials, alerting services, and customisable resource finders. The AHDS archives significant collections of electronic texts, databases, images and mixed media resources, and provides access to information about them, and about similar resources, located and managed elsewhere. However, the AHDS does not generally supply access to resources beyond those collected from within the research community.

They each **presuppose knowledge** of what the user requires. There is equally an assumption that the user clearly understands the differences between what they each offer. The elaboration of the services offered is based on limited user-requirements analysis which is out-of date, specific to one provider, and generally not based on research into user-needs in the light of recent technological developments. User-requirements analysis is a fundamental part of HCI [human and computer interaction] informatics. It seeks to design the specification of ICT-ware with a real understanding of the people who use the technology, resulting in more effective tools, work practices and more successful outcomes. Its techniques are developed from social-science methodologies and vary in the amount and depth of information to be obtained and the level of intrusiveness to the user. That analysis was not available for this service area. The RePAH Project was established to provide it.

Alongside the elaboration of these complementary service-providers there has also been a rapid development in **new 'pervasive' technologies** that refine, personalise and render interactive subject gateways and portals (through tool-bar type tools or portlet developments). An essential part of the background to the RePAH Project was therefore

to examine current information search/access strategies and patterns among research practitioners and develop 'demonstrators' to investigate future user requirements for advanced information services that will serve to facilitate greater take and up use of these resources.

3.2 Aims and Objectives

RePAH has the following aims:

1. To **analyse** what user-requirements analysis has been undertaken in the past to define the strategic development of portals in this area, specifically with reference to the RDN and AHDS.

2. To **survey** current user-needs, as defined by their information search and access strategies and patterns by arts.

3. To **identify** the future needs in the UK arts and humanities research communities for the development of more refined, personalisable, interactive, integrated portal services ['portlets'].

RePAH's overall objective is to make recommendations on the basis of the above for the **further development and possible cross-linking** of these services, based on a sound understanding of user-behaviour, requirements and preferences.

3.3 Definitions used in this Report

There is no agreement in the literature on what the term 'portal' means. That has not, however, stopped its being frequently used in the context of environments of networked information. In reality, the term is used within a **spectrum of meanings** that reflect one or more of the following distinct, but complementary functionalities:

- o An IL [information location] that **links** distributed sites of information [manually-harvested ↔ mechanically-harvested LINKS]

- o An IL that **evaluates** sites of information [searchable resource descriptors ↔ customised resource descriptors to particular individuals/needs]

- o An IL that **federates** distributed sites of information, encoded with metadata [structured metadata in forms and search-results that are readily understood by the user ↔ structured metadata where the user needs to be assisted in understanding the origin, form and results of the data]

- o An IL that **orchestrates** network search environments and applications to provide additional or personalised information for the user [multiple functionalities based on cross-searching or metasearch ↔ portlet applications, personalised access, processing and delivery of such information]

o   An IL that **manages** access to networked information on a predetermined basis [managed access within an organization/institution ↔ managed access to information outside an organization/institution that has been commercialised or otherwise protected]

On the basis of these functional spectra, which are not exclusive to one another, we propose to assign the following meaning to the terminology we adopt in this report:

a)   **Gateway:** A gateway places the emphasis on providing **links** to distributed sites of information. A gateway service may also evaluate the resources enumerated. Within the RDN context the services provided by a hub, an organisational entity comparable to a subject centre.

b)   **Portal:** a portal places the emphasis on **federating** distributed sites of information. This conforms to the JISC definition:

> *Technically, a portal is a network service that brings together content from diverse distributed resources using technologies such as cross searching, harvesting, and alerting, and collate this into an amalgamated form for presentation to the user. This presentation is usually via a web browser, though other means are also possible. For users, a portal is a, possibly personalised, common point of access where searching can be carried out across one or more than one resource and the amalgamated results viewed. Information may also be presented via other means, for example, alerting services and conference listings or links to e-prints and learning materials.* (JISC 2003e)

c)   **Managed Information Environment:** a managed information environment places the emphasis on **managing access** to information, structured for the use of those within that environment. These employ 'portlet'-style technology to provide additional or personalised information services for the user.

3.4 Methodology

3.4.1 Stakeholder analysis

Our research process began with discussion and definition of the stakeholder groups for this study [Appendix A2]. We identify these as:
o   Researchers
o   Service providers
o   Funding bodies

The main target user group, 'Researchers' was further refined as:
o   Postgraduate [PG]
o   Postdoctoral [PD]
o   Research Assistant [RA]
o   Faculty
o   Independent researchers [IR]

3.4.2 Research questions

The project aimed to collect four different kinds of data related to research portal needs:

1. Information about users' awareness and attitudes with respect to currently available online services and tools including such gateways and portals as currently exist.
2. Information about user's information discovery strategies and internet usage.
3. Information about patterns of recent user-activity in relation to the RDN subject hubs and AHDS.
4. Information about users' responses to what future portal developments can deliver.

3.4.3 Research methods

The research methodology adopted draws on two complementary paradigms. Firstly, the main thrust of our investigation was historical and evaluative, that is to say it aimed to 'discern patterns of use and to collect qualitative statements regarding the use and improvement of the various [….] components'. In broad terms this approach can be situated within the **design-based research paradigm**. Design-based research is carried out in a continuing cycle of design, enactment, analysis and redesign. Within this study we have picked up the cycle at the enactment stage, conducted an analysis of the current picture and used the redesign stage to explore user-reactions to possible future functionality through prototype demonstrators. Secondly, however, the focus on primarily qualitative data about peoples' behaviour and attitudes situates this study also within the domain of **applied social-science research**. Within these two broad frameworks a **mixed-method approach** was adopted, combing quantitative and qualitative techniques to achieve the best results in terms of addressing the information types required and allowing the possibility of triangulation of different data types.

3.5 Data Sources

Data was provided as follows:
- o   Published Reports and Evaluations of Service Providers
- o   Questionnaire survey
- o   Focus Groups
- o   Delphi
- o   Server log analysis
- o   User trials

3.5.1 Published Reports and Evaluations [see Appendix A3]

We examined all available Annual Reports of the AHDS and its constituent branches, as well as the two RDN 'hubs'/'portals' in the period since their creation. We paid particular attention to any user-evaluation work that was undertaken. The more detailed evaluation of this evidence is considered in **Appendix A3.**

3.5.2 The Questionnaire [see Appendix A4]

Survey research aims to measure certain attitudes and/or behaviours of a population or a sample, most often by asking respondents for information. The survey instrument used was an online questionnaire on the project website, linked to from a number of related sites, in particular AHDS and Humbul. Potential respondents were alerted to the questionnaire through links embedded in these websites, plus email lists, newsletters of professional associations, online community websites and journals.

3.5.3 Focus Groups [see Appendix A6]

Focus groups combine elements of two other social-science research methods: interviewing and participant observation. The advantage of focus groups over interviewing is the explicit use of the group interaction to generate data and insights that would be unlikely to emerge without the interaction found in a group. An important aspect of conducting focus groups is the topic guide. The topic guide, a list of topics or question areas, serves as a summary statement of the issues and objectives to be covered by the focus group. It also provides the initial outline for the report of findings. The topic guides and evidence from the focus groups is presented in detail in **Appendix A6**. To conform with data protection legislation, the transcripts of the focus groups will not be archived with the rest of the project. The first round of focus groups addressed research questions 1 and 2. The second round was used as part of the iterative process to gauge user-reactions to different future scenarios of portal development by discussing a prototype 'demonstrator' portal, discussed in detail in **Appendix A8.**

3.5.4 Delphi [see Appendix A7]

The Delphi technique is a systematic, iterative, predictive research method based on independent inputs from a panel of experts. The objective of most Delphi applications is the reliable and creative exploration of ideas or the production of suitable information for decision making. Delphi is based on a structured process for collecting and distilling knowledge from a group of experts by means of a series of questionnaires interspersed with controlled opinion feedback. It measures the degree of consensus among the panel regarding future events where the decisive factors are subjective, and not knowledge-based. The technique reaps the benefits of group decision making while insulating the process from the limitations of group or peer pressure and overly dominant individuals. The technique involves iterative rounds of questionnaires where responses are re-circulated so individuals can reconsider their opinions in the light of the responses of the panel as a whole. Our Delphi Exercise on Portals is further examined in **Appendix A7.**

3.5.5 Web server log analysis [see Appendix A5]

Web server logs record simple traffic statistics and data such as the numbers of page requests per month and originating addresses of page requests. Deep log analysis (DLA) uses web logs from a server and after the normal process of analysis links the information with site user profiles, or demographics, to produce a 'deeper, more meaningful data' picture of overall site usage. It is a four stage process:

o **Data definition** where recording the procedure and statistical significances are agreed.
o A series of **pre-defined metrics** are used to ensure the data is analysed in line with organisational goals and policies.
o Enrichment of usage data with **demographic data**.
o Identification of **questions concerning information-seeking behaviour** that need to be clarified by other user investigation.

An example of working metric definitions are:
o **User**. A user is effectively a computer; sometimes that computer represents an individual, in other cases a number of people. User identification can be based on a combination of 'IP' number and browser details, or by use of cookies.
o **Sessions**. They are identified in the logs by a session identification number. Logs include a session beginning tag and a session ending tag, which enables time calculations as well.
o **Items viewed/requests made**. The key usage sub-metrics are: type of items viewed, number of items viewed in a session and return visits. These sub-metrics offer good platforms for characterising and comparing the information-seeking behaviour of sub-groups of users.

A more powerful way of examining the number of items viewed is to categorise search sessions by the number of items viewed. This is called **'site penetration'**. Research on the subject has shown that many web users graze lightly, examining just a few items/pages before they leave with no substantial content consumed, although knowledge might have been gained. High levels of penetration can be assumed when there is evidence of:
o 'natural movement' through the site
o the investigative nature of information-seeking
o the presence of an embedded search engine and other retrieval aids
o return visits to a site.

3.6 Problems with the Data

We have taken into account the following deficiencies in our data:
o **Incompleteness.** Annual reports are not available for all the services since their creation. Some user-evaluation undertaken in-house was not published. The evidence from web-logs was not archived for one of our services (Artifact) for the period of a full year.
o **Unrepresentativity.** The target population for our population was too large for us to survey comprehensively. We adopted a sampling approach in our focus groups. A non-probability sampling approach was used (self-selected sampling) in which the respondents chose whether to be included in the survey. Although less reliable than simple random sampling, stratified random sampling, or proportionate sampling, where care is taken to ensure that the sample is not biased in some way, this was the only option available to the project. The responses may not, therefore, be fully representative of the population as a whole. In particular, it

is reasonable to suppose that the respondents are biased to some degree *in favour* of ICT since (a) the subject of the survey was the use of ICT in research and (b) the survey questionnaire itself was itself available only via the Web.

o **Comparability.** Our evidence was not always comparable. The methods of presenting usage data in published Annual Reports are not comparable with one another, and often on unclear bases.

o **Disaggregation.** Because of the complexities of the server-structure within the AHDS, we do not believe that our web-log analysis covered all the site activity at all the sites. It proved impossible to strip out the 'internal' AHDS log referrals in a way that satisfactorily disaggregated site consultation from other traffic.

o **Interpretation.** Deep-log analysis, in particular, poses problems of interpretation of the evidence it affords. Although it is based on what can seem very impressive samples, these can camouflage substantial differences between individual user groups. It enables us to map the digital environment of the service providers more accurately but it cannot, on its own, provide much by way of explanation, levels of satisfaction recorded, and the impact of the consultation upon the user.

For these reasons, this report is based on a triangulation approach, looking for the reinforcement of the evidence from one set of data in another before drawing strong conclusions on the basis of it.

4. The Arts and Humanities Research Community

4.1. Subject-Domain Analysis

What is the profile of the 'arts and humanities research community'? Our analysis is limited to an answer to the question that is sufficient to understanding its resource discovery needs. We understand it as **non-homogeneous, institutionally diverse,** and **variegated** in its research patterns. In comparison with other scientific disciplines, however, it has some **distinctive cultural approaches** that affect the way in which it approaches its resource discovery needs.

4.1.1 How many disciplines make up the 'arts and humanities research community'?

We have taken the eight panel profile of the AHRC, and mapped onto it the RAE subject panels. We have then compared these with the subject coverage of the RDN portals and AHDS service providers **[A2.1]**. The **subject breadth** of the community needs to be emphasised, since it underlines all the difficulties of subject-specific resource discovery in this area. Twenty-five of the 69 2001 RAE Panels fall within this area. Many of the subject areas are small and discreet. Many research practitioners would probably not regard themselves as part of anything as coherent as an 'arts and humanities research community'. The current electronic resource-discovery aids do not provide an 'even provision' to the disciplines in question. The least well-served area is Panel 8 [Philosophy, Law, and Religious Studies]. It is no coincidence that this was the area where we had the greatest difficulty in establishing a reliable user-requirement response. Other areas are clearly only partially served by the current providers. At no stage in our investigation, however, was any comment made to us about this unevenness of provision. **We registered no strong sense of perceived comparable inadequacy from practitioners in any particular disciplinary area.**

4.1.2 How many are involved in 'arts and humanities research'?

We do not know the answer to this question. Working on the basis of our initial stakeholder analysis (**3.4.1**) we used the RAE2001 returns to provide us with an overall pattern of **research population** by AHRC Subject Panel [**A2.2**]. Just over 12,750 practitioners were recorded as research-active in that review. Although the statistics are now five years out of date, we do not believe that the overall pattern will have greatly changed.

Of course, these figures do not take into account the other stakeholders that we have identified. In the case of Arts and Humanities **PGT and PGR**, we have used the statistics for HE qualifications obtained in the UK for 2004-5. These are broken down into very broad categories. Just over 27,000 individuals successfully completed their

degrees in the Arts and Humanities during that period. Applying appropriate year-cohort multipliers, this would indicate a **PGT and PGR** demographic of 35-40,000 [**A2.4**].

Arts and Humanities **RA** [Research Assistants] and **IR** [Independent Researchers] are categories that elude us. There is likely to be a varying penumbra of users according to the discipline in question. The Archaeology portals, for example, will be used by UK archaeologists from a variety of backgrounds, of which only a small proportion (perhaps around 15%) are within HEI [**A2.5**]. This is probably an exceptional case. We imagine that a multiplier in the range 10/50% of HEI established posts is a reasonable working hypothesis (i.e. between 1,275 and 6,375). According to Abbott and Beer (2006) there are some 30,000 employed in the music, visual and performing arts sector.

Of course, the RDN portals were designed to be used for both teaching and research purposes. They have a wider remit than simply for conducting research. Our effort has not been to calculate the overall demand for portal services, simply that part of it which we might define as (at least in part) driven by a research agenda. For the purposes of this report, our assumption is that a reasonable estimate of the per-annum demographic cohort for Arts and Humanities portal services is of the order of **c.50,000-60,000**.

4.1.3 How are they scattered?

We further categorised the HEI return for each subject area in the RAE2001 exercise by size in order to achieve a picture of the subject distribution. It confirms what is generally known. Research-active Arts and Humanities units are generally small (under 10) to medium-sized (under 30). Only a small minority of units were recorded as large (over 30) or very large (over 50) [**A2.2**]. This pattern necessarily has an impact on the research information needs of users, making distributed information generally more significant as a way of keeping abreast in particular research fields.

4.1.4 Mapping Arts and Humanities Research Activities

Arts and Humanities research is variegated in nature. In pursuing the research for this project, we needed a **road-map** to understand it better [**A2.4**]. The diagram emphasizes that:
   o Arts and Humanities researchers have, in addition to their **core** research activities, a **penumbra** of research-related activities, for which there are significant information resource-discovery needs.
   o Arts and Humanities researchers have both **individual** and **group-based** research activities
   o Arts and Humanities researchers are not universally served by the current information resource-discovery channels in all these areas

4.2 The Characteristics of Arts and Humanities Research

How should we characterize the fundamental characteristics of Arts and Humanities research? The question takes us well outside the brief of this project. But some appreciation of **disciplinary difference** is important because there is a risk that models of scientific activity derived from the pure or applied sciences are applied inappropriately and that, as a consequence, information-discovery tools are not fit for purpose.

We have understood the differences between disciplines in terms of their **knowledge structures** and their **cultural characteristics**. The resulting model, outlined in a classic formulation by Becher (1989) and summarised by Fry (2004), emphasizes the following fundamental characteristics of Arts and Humanities [**A2.7**]:

- o **Reiterative knowledge-gathering processes**. These are typically **open-ended**. They do not depend on **clearly-defined taxonomies**. They are suspicious of **categorised information**, preferring often to deal with particulars, qualities and complication. They often prefer to undertake their own **ontological evaluation** of knowledge.
- o **Individualistic and pluralistic cultures**. The research agendas are defined at an individual level and the communication networks are localised, extensive and informal. The research agendas are often not well-understood beyond the particular individual in question. The research is loosely structured. There is sometimes an underlying 'counter-culture' which is suspicious of conformism and authority.

It should be emphasised that these are models, and not stereotypes. All patterns have exceptions, and these descriptors can readily be challenged with counter-examples. That said, these are models that we have found useful in our analysis, because they have led us to some **fundamental features** underlying Arts and Humanities research behaviour as regards resource discovery and information needs:

- o All machine-accessible resource discovery depends upon implied taxonomies of knowledge. Pre-structured knowledge is not always greatly appreciated. The need to be assured about the authority of a particular digital resource is balanced by the assumption that the individual researcher has specialist skills with which to assess its authority. Arts and Humanities researchers are as likely to want to develop their own resource discovery trajectories as to follow those dictated by others.
- o Categorised information is often not ideal when one is interested in the particular, or the anomalous. Resource discovery can provide pointers in the right direction, but Arts and Humanities researchers readily accept that individual resource discovery is fundamental to their research. The reiterative processes that this involves are a key constituent in the pursuit and definition, of their research agendas.

o   Arts and Humanities research is still mainly defined at an individual level. Information resource development has to be based upon these individual needs.

o   There is likely to be a significant relationship between the relative lack of 'mutual dependence' among Arts and Humanities researchers, their 'task uncertainty' and the ways in which digital resources are produced and utilised. Although the documented evidence for this is based on exemplar fields that are not specifically within the Arts and Humanities, the trajectories are applicable [**A2.8**]. The arrangements for collaborative research and disseminating research results are personalised, localised and decentralised. Informal communication depends on individual groups and specific social networks. Digital resources, where they exist, tend to be field-based and similarly localised. Alternatively, there is a reliance on commercially produced generalist digital resources.

o   Arts and Humanities 'e-infrastructure', apart from the service-providers that we are investigating, tends therefore to be determined at the level of the employing institution rather than the field or discipline, or higher.

o   We have to take into account a 'counter-culture' (which we encountered most noticeably in our investigation of some of the disciplines in the Arts, where information resource needs and research agendas are often articulated in terms of 'diametrical difference' to prevailing trends. Artists predominantly wish to be known for distinctive differences and not part of the crowd, and any associations with an established authority risked being influenced too much by dominant trends.'

o   Arts and Humanities researchers want access to information, irrespective of the medium in which it is available. They are used to working in fields where there is a very mixed economy of resources, electronic and physical. Journal articles are important, but so are printed books. E-prints (pre- and post-) are markedly less significant than in the physical sciences and engineering. Electronic bibliographical information is therefore of critical importance to Arts and Humanities researchers. Within the UK's Higher Education Institutions there is a growing movement to develop institutional repositories. Those being established are within the sciences and social sciences, but as of yet they have not developed as vigorously within the arts and humanities.

o   There is less emphasis on communicating work-in-progress and more emphasis on formal ways of disseminating information. There is consequentially less emphasis on lead-times for accessing research-sensitive information and results.

4.3 Conclusion: Arts and Humanities Information-Seeking Behaviour

Arts and Humanities scholars want access to information irrespective of the media in which it is available. They expect a good deal of that information to be available digitally, and to incorporate that into all area of their research work. There are **high and growing levels of expectation** as to the availability of materials in digital form. Those expectations are being fed by the **exponential growth** in the content of Arts and Humanities digital libraries by a wide variety of different content creators and contractors. We have no estimates of the current scale of digital libraries and content in the Arts and Humanities but it is clear that much of it is being generated outside

established scholarly and research channels. The **need** for **quality resource discovery tools** is, therefore, higher than it has ever been before. Equally, the **need** for **quality assurance** of what is provided is, also, higher than it has ever been before.

Our research has reinforced the conclusions of an earlier, broader survey in 2005 as to the information-seeking behaviour of Arts and Humanities scholars [**A2.8**]:

- The **resources** that they most seek access to are: books, articles and non-textual materials, in particular digital image collections.
- The **search tools** that they most use to find these resources are: search-engines, bibliographic resources; and, subject-specific abstracts and indexes. Subject-specific portals are not currently a way by which many Arts and Humanities scholars find their resources. They use works of reference more frequently than they consult such gateways.
- The **informal resources** that they use include: emailing colleagues, asking colleagues, reading email newsletters and posting enquiries on email lists and bulletin boards.
- The **problems** that they encounter in accessing resources are dominated by the following: the particular HEI does not take the books/journals/subscribe to the databases the individual researcher needs; the need to travel to access resources which are either not available in digital forms, or not distributed digitally.
- Key research information is only available in proprietary digital media. This is a particularly significant problem in some areas of the Arts and Humanities research domain, especially in the Visual Arts (film, photography, art) and some large historical datasets.
- Overall the **perception of problems** in accessing resources do not appear to be significantly greater (overall) from those in other disciplines.
- **Habit and familiarity** play a large part in information-seeking behaviour. There is a recognised trade-off between the amount of time a user is prepared to spend in learning about an information resource tool, and their choice to use it. Arts and Humanities Scholars make the equivalent of a personal cost-benefit analysis when it comes to being prepared to use a particular resource-discovery tool, reflecting a differential sum of the following elements:
  - speed and proficiency
  - transparency of results
  - perceived relevance, density and completeness of the resources discovered
  - authority of the results recovered
  - ability to manipulate (download; transfer) the results
  - recommendation from others as to its utility
- Because of the dispersed nature of the disciplines involved, formal programmes for training and familiarization in the Arts and Humanities are difficult to deliver and rarely cost-effective. On-line tutorials for resource discovery have been extensively developed by the RDN network. Our users have not, however, made significant use of them. In reality, most users concentrate around a limited number of frequently-consulted resource discovery tools – sometimes as few as 4-5.

- o The relative agility of the informal networks of communication in the Arts and Humanities means that there is a good deal of lateral 'shared knowledge' within disciplines about what resource discovery tools are most fit for purpose.

5 User-Requirement Analysis for Portals in the Arts and Humanities

5.1 The Documentation

The bulk of the Project's work involved the undertaking of this user-requirement analysis. This involved:

- o a thorough understanding of the evolution of the current service provision on the basis of their published literature [**A3**].
- o a triangulation of user-needs analyses, using the established techniques of the applied social sciences. These included a first round of 'focus groups' and interviews [**A6**], an online questionnaire [**A4**], and an analysis of the web-log data of the current service providers [**A5**].
- o an iterative process, enabling users to articulate their needs. This involved a Delphi exercise [**A7**], the development of a set of mock-up demonstrators of potential portal developments [**A8**] and a final set of user-trials of these developments [**A9**].

This part of our report provides an overview of the detailed findings in these individual work-packages and reports.

5.2 The Arts and Humanities Portal Problem

It will be helpful here, before entering into the more detailed issues which emerge from these reports, to outline what the underlying 'Arts and Humanities Portal Problem' appears from this evidence to be. At the risk of over-simplification, we present it as a series of propositions:

- o Arts and Humanities scholars need access to a very wide diversity of research materials in digital media, growing very rapidly, furnished by a variety of commercial and non-commercial providers, in different formats and standards, often addressing different disciplinary needs and agendas, maintained by different bodies, only some of which are UK-based.
- o The current subject portal and digital archive repositories do not provide access to the majority of these materials. They do not harvest the metadata from them. They do not provide interoperability. Nor do the other institutional portals.
- o There is therefore a mismatch; users have diverse resource-discovery needs, which the resource discoverers do not, in the main, satisfy.
- o There are understandable reasons for this mismatch. They include the following
    - — the RDN subject portals are locked into an out-dated methodology of manual harvesting and resource authentication.

- the AHDS concentrates on the archiving of digital materials, but does not archive their functionality. They thus become fossilised deposits for the user, relatively inaccessible. Again this is partly due to the mission of the AHDS to simply collect from the research community and not for it.
- the providers have not developed a coherent strategy for understanding evolving user-needs
- rapid technical evolution has meant the swift emergence of new technologies for individually managed information resources within more collaborative frameworks.

o The 'portal problem' has been 'latent' in the Arts and Humanities because more sophisticated commercial internet search engines have answered some of the immediate needs. But these search engines are relatively inflexible. Users are not convinced by their ranking systems. They are unsure of the authenticity of the information they provide and overwhelmed by its inherent redundancy. Increasingly, Arts and Humanities users are becoming aware of the problems that these internet search engines do not address: access to online digital resources which have not been opened to harvesting by search engines; and the related lack of interoperability between digital libraries, each hermetically sealed from one another.

o Yet emerging technologies do provide potential solutions to this mismatch problem. With emerging metadata standards, there are greater possibilities for automatic harvesting techniques. With better desk-top tools, there is more opportunity for the personal management of resource discovery. With different management of digital resources, functionality can be maintained along with datasets. The research communities can themselves be more involved in the provision of metadata for digital resources, and in authenticating them.

o This depends on a mediated environment through the widespread adoption of a common authentication system.

o The overall objective is to create a managed digital research environment in which access to resources is increased, alongside a greater interactive functionality in relation to them. The possibility for a greater array of scholarly communication needs to remain under active consideration.

5.3 Patterns of Arts and Humanities Digital Research

Our investigation confirms the ubiquity of digital resources for Arts and Humanities research. Over 60% of respondents to our online questionnaire regarded digital resources as 'essential' to their research [**A4**]. These resources were used 'extensively' by a majority of our respondents. Digital resources were emphatically not restricted to teaching delivery. In fact, whereas only a minority of our online questionnaire respondents thought it had changed the way that they taught, a clear majority thought that it had altered the way that they undertook their research. The first set of focus groups reinforced that sense – emphasizing that the existence of digital resources had changed

the way that their agendas for Arts and Humanities research had been formulated, as well as transforming the way in which the material for answering those research questions was discovered and analysed [**A6.2.1**]. At almost every stage of the research process, digital resources have changed the way that Arts and Humanities research is now conducted.

We should nuance that conclusion in three important respects:
> 1) Our first focus group and our questionnaire evidence suggests that it has not yet profoundly influenced the way in which Arts and Humanities **publication** is conceived, where digital publication is not yet perceived as a logical consequence of the changes to research processes [**A6.2.1**].
> 2) This change in research process has not fed through to the habits and procedures for **personal digital data archiving**, where (according to the evidence from our online questionnaire) our respondents are not particularly engaged by the issues [**A4.3**].
> 3) This change has not yet had a substantial impact on the means **of scholarly communication** in the Arts and Humanities. The evidence from our first set of focus groups and questionnaire responses was here confirmed by the lukewarm reactions to the possibilities for more elaborate forms of online scholarly communication that we discussed with them. The sophisticated, lateral research networks in the Arts and Humanities seem adequately served by the current range of email, bulletin boards, and blogs (only occasionally used for research purposes, according to our research) [**A4.5; A6.2.3**].

Our respondents emphasised that their research agendas were flexible, open-ended, established on a predominantly personal basis, and multiple. Thirty-one percent of our online questionnaire respondents regarded themselves as not having one single research domain. Eighteen percent said that they had several. General scholarship was regarded as central to over 60% of our respondents' work. They did not regard themselves as working in hermetically-sealed specialist areas. Rather, they saw themselves as researching overlapping domains, in which there were a series of core issues which could be tackled from a variety of differing angles. They accepted that there was a distinction between 'core' and 'penumbra' research, although they wanted to keep many aspects of the 'penumbra' of research (e.g. refereeing articles for a journal; refereeing research proposals, etc) at arms' length wherever possible [**A2.9**].

What digital resources did our users find most useful? How did they find them? Here, we were impressed with the very broad range of digital libraries, bibliographical tools, encyclopaedia, dictionaries, and other online materials indicated to us by the respondents to our online questionnaire [**A4**]. Our focus-group participants were anxious to reinforce the message that they were practical-minded and instrumental in what they used, concerned about access to them, and resourceful in the way in which they searched for more materials of relevance to their work [**A6.2.2**]. The patterns were quite discipline-specific. Their needs were extensive, and often indeed broad-ranging. The range of research questions was very wide. Our questionnaire respondents referred to their extensive online bookmarked resources. Our focus groups reflected researchers who expected their research methodology to involve a high degree of proficiency in resource

discovery. They learnt about resources from other practitioners by lateral means of communication. Their levels of formal initiation or training in the digital resources that they used varied from little to none.

The range of service-providers for resource discovery was correspondingly varied. University Library services and catalogues (OPACs: COPAC) are evidently significant. Internet search engines are regularly used. Users are not promiscuous, but they have formed their views on the perceived cost-benefits of using particular resource-discovery strategies for their purposes. Those views are necessarily framed on sometimes a less-than-perfect appreciation of the possibilities and range of a particular resource discovery tool or digital library.

We were particularly interested to discover the impact of the RDN subject portals and the AHDS as resource discovery tools in this pluralist environment. We first studied the evolution of these two services since their inception about a decade ago [**A3**]. We noted a degree of patchiness in the coverage of the Arts and Humanities disciplines [**A3**]. The services themselves had evolved independently of one another, although they had complementary missions in the resource discovery area. Although the pattern varied across the services, there is no coherent strategy for consulting users and discovering their needs. After a decade of development, the services are not interacting well with one another at the resource-discovery level. So, while the AHDS and the RDN (Intute) have their collection metadata in OAI (Open Archive Initiative) formats, available for harvesting, it is not picked up comprehensively by the either of them. Equally, although Intute was launched in July 2006, at the time of writing [September 2006], there is almost no mention of its existence on the AHDS site. Because of the breadth of Arts and Humanities digital resource needs, and the diversity of their information providers, resource discovery services, tools and mechanisms need to be based on a strong collaborative framework, engaging with the major research libraries, archives and other creators and holders of digital content. The AHDS has developed important links in individual subject domains. The RDN, however, appears to be more limited in its collaborative frameworks.

Our users were clear about the potential importance of authenticating digital resources, although they were not so sure about the resource descriptions in the RDN subject-portals. In particular, they had no sense as to how often they were up-dated, and what range they covered. Those that had used the subject portals, took the view that they tended to be useful at the beginning of a research enquiry, but to become progressively less relevant as it deepened. As for the AHDS, although the number of its resources downloaded seems to be increasing, none of the participants in our focus groups or questionnaire admitted to having downloaded such collections. Where the AHDS harvested data, generally in collaboration with outside partners (as in e.g. Heirport the Historical Environment Information Resources Portal), it plays a significant, perhaps pivotal, role in particular agendas of Arts and Humanities research.

So our evidence is unambiguous about the relative insignificance of the RDN portals and AHDS for most research purposes for the Arts and Humanities practitioner. Only 4%

singled them out as relevant to their digital resource needs and resources on our online questionnaire. Our web-log analysis tends to confirm that order of percentage for UK researchers as regards the site usage for Humbul and AHDS [**A5**]. (For Artifact, we had only fragmentary statistics to rely on, and the service has been in existence for a shorter period, with less time to build up its collections.) The evidence for Humbul 'site penetration' by users is more ambiguous. Academic users certainly tended to spend more time at the site than other users, and be more determined in their browse strategies. But users in general tended to come to RDN resources from external search engines than from an internal search of the site. This may reflect the fact, noted in our appendix that Humbul's OAI metadata was offered for harvesting by Yahoo where its hits rank high in search returns. It might also, however, suggest that users were consulting Humbul as part of a broader online search for materials. Although a significant proportion of those we have identified as these academic users went on to consult the summary description of a digital resource, only a small minority of the users tended to go through to link to it.

The AHDS was equally classified by our online questionnaire respondents as one of several resource discovery channels, alongside 'news and media' and the 'Web of Knowledge', of about equivalent relevance to their resource discovery needs as the RDN portals (4% of our online questionnaire respondents)[**A4.3**]. The web-log data for the AHDS suggested a rather smaller percentage of site-usage for UK researchers than for Humbul. But there are some serious potential problems with these statistics. We are not sure of the extent to which individual AHDS sites were visited separately from the AHDS server, and whether this is recorded in its web-logs. Equally, we are not convinced that the internal traffic of the AHDS within its distributed hub-structure, has been adequately stripped out from our web-log data. Our focus groups, and associated analysis undertaken for a separate review of the AHDS service [Brown et al, 2006], emphasised that the users of the AHDS included several disparate groups, with different and non-complementary needs. Users reported that the resources they found via the AHDS were often not relevant to their needs, being either too niche or too generalised, the result of collecting small, disparate data sets, with large gaps within and between subjects. As with the RDN, there is a problem of 'critical mass', an essential prerequisite to the success of a resource-discovery tool. Our users retained, however, a positive view of the AHDS, even if they do not use it much. They appreciate its role in other areas, but simply have alternative ways of meeting their digital resource discovery needs that suit them better, or which they know better.

The internet search engine emerges from this study as an immensely useful digital resource discovery tool. In certain disciplines (Classics, Ancient History, Visual Arts and Media), Google was cited by our questionnaire respondents as their central tool for acquiring digital information. And, even though our web-log data revealed that our users deployed a variety of proprietary search engines, their simplicity and speed appealed to our users, for whom a key determinant in their cost-benefit analysis of resource discovery tools was whether it saved, rather than cost them time. That said, our users were also often aware of the limitations of their internet search engine of choice. Our users told us of their frustration at its lack of sophistication (a frustration that is, we concede, often a function of their lack of familiarity, or perhaps understanding, of Boolean search

parameters permitted in Google's advanced search facilities). They were suspicious of the ranking of the hits returned, but were equally overwhelmed by the information redundancy which accompanies search-engine retrieval on internet materials. They were, above all, concerned about the fact that search engines do not search a great deal of digital content that is relevant to their needs; and, equally, they are frustrated by the lack of interoperability between different libraries of digital content.

The issue of 'access' runs through all our enquiries. Access to online journals was emphasised in the first focus groups, and reinforced in the online questionnaire and in our Delphi analysis, where it consistently came top of the list of user-needs [**A4.3; A5.7; A6.2.5**]. But the issue of access was also raised in respect of proprietary digital content of various kinds, specific to particular disciplines. The issue was sometimes presented in terms of an implied trade-off in resource terms, with our users wanting to see the investment of scarce resources in widening the local access to digital content through licence and content purchase rather than increased investment in resource discovery. At the same time, our focus group research practitioners were also aware that 'access' to digital content was not a simple matter of 'Oliver asks for more'. Access was only fully beneficial to the user when it was linked to enhanced resource discovery, and particularly interoperability.

Interoperability was another major theme running through our enquiries. It tended to affect some disciplines more than others; but it was present at some level for them all. The problem is evident to many practitioners. As digital content becomes richer and more diverse, so the independent platforms on which it is consulted multiply. As interoperability becomes more important, so the potential for a next-generation resource-discovery portal grows. But the current providers do not harvest a great deal of content. Our users, in so far as they were familiar with the RDN subject portals, were very unclear about what data, if any, they harvested. By contrast, they understood very clearly the scope and range of the COPAC catalogue and other metadata harvesters in their particular subject-domain.

Another important issue raised in the course of our investigations was that of resource authority and quality control. Our users wanted to have assurances of quality. This emerged in the first focus groups [A6.2.4]. It was reinforced in the cycle of Delphi forecasting. But they also remained suspicious about who was undertaking the quality assurance. They wanted to have a role in the process, rather than have it mediated to them.

There were a number of other issues that our users raised. In Music and the Performing Arts, there were specific technical issues about retrieving and downloading very large files, and having the software with which to consult and manipulate them. In the Visual Arts, there were specific issues around digital images, many of which echoed the recent report on the subject from AHDS Visual Arts [AHDS Visual Arts, 2005a], where issues of access and interoperability are particularly acute. The question of digitally archiving functionality with content was raised in several of our enquiries, even if our users were not fully aware of the costs and difficulties of doing so. Questions of copyright and the

use of digital content, and how to cite it, were also touched on as among the issues in our users' minds.

5.4 Portal Futures

The main thrust of the second half of our study was to investigate what features of the emerging ICT technologies for advanced resource discovery and communication would be most likely to meet the research strategies of the Arts and Humanities practitioner. In the most general terms, we identified these emerging technologies as providing tools for resource discovery, workflow management and communication. We concentrated on the greater possibilities for desktop interoperability, for more personalised management of resource discovery needs, and for the involvement of the research community in the provision of metadata for digital resources and for their authentication. The methodology in this second phase was adjusted to obtain formative evaluation feedback. It is now standard practice in product design and development that user-testing involves an iterative process of refinement and modification to adjust product development to meet user needs. The design of the Arts and Humanities research tools of the future should be no different. This was the purpose of our Delphi exercise and our final phase of user trials of portal demonstrators [**A7; A8**].

The results of the Delphi exercise [**A7**] were combined with the outcomes of the interviews, focus groups and questionnaire results to generate a list of desiderata. From these, a series of wireframe graphical mock-ups were created for evaluation purposes.

The shortlist of requirements that emerged from the earlier engagements with users was:

1. Ability to conduct simple searches across disparate data collections.
2. Ability to share ongoing research work, notes and ideas with research collaborators.
3. Ability to publicise and disseminate completed work, and comment upon other such work completed by peers.
4. Ability for comments / reviews / peer-moderation to influence searches by flagging up content that has been deemed legitimate.
5. Ability to browse through disparate resources as well as search.
6. Moderation, submission and creation of content by community as opposed to central authority.
7. Inclusion of news feeds and current event information.
8. Ability to create new searches within the context of existing searches.
9. Inclusion of background information about the creator of a piece of content, which would allow the user to assess their 'point of view'.
10. Inclusion of IPR and copyright information about resources.
11. Tracking of the user's use of resources discovered via the portal.

NB the requirement to access a wider range/all online journal content was not explicitly included in our requirement analysis, since the issue is one of content rather than functionality. But access to journals is subsumed within requirements 5, 6, and 8 above.

The demonstrators were designed to be modular in nature to allow for their extension and personalisation. They do not cover all the potential functionality, but they provide a mock-up of what a managed, customizable, portal research environment might look like. Our mock-ups focused upon the following features:

- o **The system homepage:** what the researcher would see when they logged on using their Shibboleth or other user authenticated account.
- o A typical **set of search results** that the user would see after conducting a Google Scholar search from within the system framework.
- o An example of an **annotated web page** that a researcher has visited.
- o An example of the **usage history** for a resource: in this case a paper in an online repository, though it could be a website, an online article, an entire journal, a dataset or a book from the library.
- o The researcher's **bookmark** management system. Again, all types of resources could be bookmarked, not just web pages.
- o The researcher's **online CV**. This would contain a short biography, their current job title and location and information about their projects (current and previous), their professional associations and a record of their publications.
- o A **project management** page showing details of the project team and linking to all shared documents generated by the project, as well as email and shared bookmarks that team members had collected.
- o A list of the researcher's collaborators or **research partners**. This page would also provide access to all the documents shared by research partners, all the email sent by and to them, and all the bookmarks they have shared, as well as links to their online CVs.

The resulting mock-ups are included in **A8**.

Which of these various potential features did our practitioners like most, and which did they find least attractive? Our final phase of user trials, detailed in **A9**, nuanced our conclusions significantly.

They were positive about the potential that the proposed resource management tools offered. But they wanted simple tools that required little or no input of time or personal engagement. They did not want tools that duplicated existing systems. They were wary of over-elaborate resource-discovery frameworks.

Workflow Management tools that give the researcher greater personal control over digital project resources, especially more evolved **bookmarking features** were identified as the most valuable. Some form of automated **copyright management system** to facilitate the growing concern with usage permission and intellectual property rights was also highly valued [**A9.3.2**].

Resource Discovery tools that provided greater control over web-based resources were highly valued by researchers [**A9.3.1**]. The ability to **filter** the quality of hit returns,

search multiple databases was at the top of all responses. Journal articles and online bibliographical resources are consistently seen as the most important and regularly consulted online resource by most arts and humanities researchers. The option to have comprehensive access to these was consistently the top request of capabilities that were proposed. However, respondents also consistently wanted these features on their terms, gaining greater control over the searching process and reticent towards the notion of contributing personal time and information to learning a new system. A **web-based news feed feature** appealed to most respondents. Respondents liked the idea of a Really Simple Syndication (RSS) style system which by-passed personal email accounts, but notified users of conferences, funding, jobs and new research publications. But they wanted these features readily customizable, so that they could be switched on and off at will, and adapted to their own specific needs and requirements.

Communication tools were not valued highly [**A9.3.3**]. Users are satisfied with existing communication systems, particularly email. Real-time 'chat' and desktop video-conferencing ranked consistently among the lowest of all tools proposed. However, **collaborative research tools** such as social bookmarking, annotating digital resources, shared document editing, attaching metadata to personally-created digital resources, and contributing to the authentication of digital content online ranked towards the middle of most responses.

Automatic information-harvesting tools were highly valued when applied to digital content to which users wanted access [**A9.3.1**]. The application of these tools to their own 'content', however, was regarded as problematic. Two automatic-harvesting tools were proposed in the demonstrator mock-ups. They proved, as we expected, to be the most challenging elements of our vision of a managed research environment. These were:

> a) an automated monitoring of electronic resource usage by research practitioners (to assist in shaping user-needs for the future)
>
> b) an automated harvesting of individual practitioner CV details to provide the basis for a national register of research practitioners and to underpin an authority system in relation to individually supplied rankings and comments on resources.

These both raised issues for our users of the potential infringement of personal privacy. They challenged the predominantly individualistic scholarly culture. There was a concern, particularly marked among early-career academics, about the possible abuse of such information.

It is worth noting that in practice it is already not difficult to create a profile of an individual from the tracks they have left in the web, nor to form a judgement about their relative standing in their field, so the concerns raised here suggest a lack of awareness about the extent to which actions are already monitored and recorded.

5.5 Summary

Our research practitioners did not want to disassociate the development of functionality from broadening access to content. Indeed, given the choice, they would prefer investment in the latter to the former. However, they accepted that the two were

intimately related, and that there was scope for additional functionality, so long as it was simple, adapted to their needs, did not replicate functionality available elsewhere, was not monolithic, was capable of being managed by them, and requiring no significant investment of time to understand and use. These are strong design constraints; and there is an implicit, but understandable incompatibility between wanting increased functionality, but not wanting to invest time and effort in understanding how it works.

Our practitioners had elaborate research resource discovery needs, and were resourceful in finding the means to meet them. The key constraint that they expressed to us was the limited interoperability. This was expressed in terms of the very limited metadata harvesting of digital resources in the Arts and Humanities, and the equally limited interoperability as between bibliographical tools and the digital resources that they catalogue.

Our users responded positively to the possibilities of a **personally-managed** research environment. There were specific, realizable functionalities that they identified as being of direct use to them in carrying forward their research agendas. These were in particular, some specific **workflow management tools** and **resource discovery tools**. Researchers wanted greater personal control over digital resources. They readily perceived the advantages of tools which enabled them to integrate searching the web with searching their own hard-drive. They saw benefits to more evolved **bookmarking** features, **personal editing** features, and an **automated copyright management system**. They wanted to be able to **filter** the quality of hit returns, **search distributed databases.** They responded positively to a **web-based news feed feature**, and liked the idea of **RSS feeds** that by-passed personal email accounts.

Our users were not sufficiently familiar with technological developments to be aware that they could play a role in adding metadata to digital content which they created so that it could be automatically harvested. Nor were they cognisant of the possible impact that their contribution could make to the authentication of online digital resources.

It is possible that, with increased IT awareness future researchers will be more tolerant of the various ways in which their online behaviour is tracked, in exchange for the enhanced resource discovery this can afford.

The tools that were intended to foster collaboration and harvest new data required that the users contribute personal data and allow monitoring from among the participating community. However there was great reticence among respondents for this degree of interaction. Anonymity and personal privacy outweighed the benefits of resource access or workflow efficiency.

They did not want additional communication tools. Automatic harvesting of their own digital content, even when it was focused on providing materials for tools that would enable them to access more readily the publications and activities of colleagues, was regarded as problematic.

6. Conclusions and Recommendations

6.1 Conclusions

This was an **information-gathering** project. Our brief was to discover **user-behaviour** and **user-needs** of researchers in the Arts and Humanities in respect of portals. We set out to discover four kinds of information:

1. Information about users' information discovery strategies and internet usage.
2. Information about users' awareness and attitudes with respect to currently available online services and tools, including such gateways and portals as current exist.
3. Information about patterns of recent user-activity in relation to the RDN subject hubs and AHDS.
4. Information about users' responses to what future portal developments can deliver.

Throughout our report, we have interpreted the concept of 'portal' within **parameters of different kinds of functionality**. They all relate, however, to **'resource discovery'**: i.e. what resource-discovery tools did researchers use most? What, in a period of rapidly-changing technical possibilities, will they want in the future?

We have gathered information from a range of sources and, applying methodologies derived from applied social-science and design-based research, allowed one element of the evidence to support and reinforce another, 'triangulating' between different data types, and being aware of the deficiencies in the relevant evidence at each stage.

Our initial analysis of the Arts and Humanities Research Community's research behaviour was substantially confirmed. This is a community which is **non-homogeneous**, **institutionally diverse** and **variegated** in its research patterns. We estimate it as around **50-60,000 active practitioners**, composed of the 'stakeholders' identified in our report – Postgraduate [PG], Postdoctoral [PD], Research Assistance [RA], Faculty and Independent Researchers [RI]. Our 'road-map' of their research activities indicated a **core** and **penumbra** of activities, which are both **individual** and **group-based** [**A3**]. Not all these activities are universally served by the current information resource-discovery channels.

6.1.1 Users' information discovery strategies and internet usage

We emphasise the following features underlying Arts and Humanities research behaviour as regards their digital resource-discovery and information needs:

o   Digital resources are now **ubiquitous** for Arts and Humanities research. They are used **extensively**. Researchers believe that they have **fundamentally altered** the way in which they undertake research – i.e. the formulation of their research questions as well as gathering materials for answering those questions. At almost every stage of the research process, digital resources have changed the way in which Arts and Humanities research is now conducted. It has not yet, however, affected the way in which Arts and Humanities **publication** is conceived (although many journal papers end up on the Web). It has not fed through to the habits and procedures for **personal data archiving** nor has it had a substantial impact on the **means of scholarly communication** in the Arts and Humanities.

o   Our researchers emphasised that their agendas were **flexible**, **open-ended**, established on a predominantly **personal** basis, and **multiple**. They did not regard themselves as working in hermetically-sealed specialist areas. Rather, they saw themselves as researching overlapping domains, in which there were a series of core issues which could be tackled from a variety of differing angles.

o   Our researchers are **practical-minded** and **instrumental** in their resource-discovery strategies. The patterns were quite **discipline-specific**. Their needs are **extensive** and **broad-ranging**, reflecting their agendas. They expect their research methodology to involve a high degree of proficiency in resource-discovery. Our users are not promiscuous, but they have formed views on the perceived cost-benefits of using particular resource-discovery tools and strategies. These views are necessarily based on a sometimes less than perfect appreciation of the possibilities and range of a particular tool or digital library and of the possibilities of ICTT generally. Both the questionnaires and focus groups highlighted a demographic within the arts and humanities community. There is a clear minority of scholars who are fluent in the use of digital applications and a sizable majority who find little need and/or time to use such tools. This finding is supported by the LAIRAH project's research which noted that there exists,

> *…a divide between the enthusiastically digital (who appear to be a minority) and the majority of the academic profession. This is worrying, since there is a danger that digital humanities may therefore become ghettoised rather than further integrated into scholarship* [Warwick, et al 2006]

o   All machine-accessible resource discovery depends upon implied taxonomies of knowledge. **Pre-structured knowledge** is not always greatly appreciated, however, by Arts and Humanities scholars. Their need for **assurance** about the authority and trustworthiness of a particular digital resource is in tension with the assumption that the **individual researcher has specialist skills** with which to assess its authority, by a **suspicion about who is undertaking the authentication**, and by an **awareness of the complexity that such a process entails**. They want to know about who has undertaken the authentication, and how often it is updated. They learn about the reliability of digital resources mostly from other practitioners, using established and informal lateral means of communication within specialist fields. Arts and Humanities researchers are as likely to want to develop their own resource discovery trajectories as to follow those dictated by others.

o Categorised information is often not ideal when one is interested, as Arts and Humanities scholars often are, in the particular, or the anomalous. Resource discovery can provide pointers in the right direction, but Arts and Humanities researchers readily accept that **individual resource discovery** is fundamental to their research. The reiterative processes that this involves are a key constituent in the pursuit of, and definition of, their research agendas. Since Arts and Humanities research is still mainly defined at an individual level, information resource tools have therefore to be based upon these individual needs.

o There seems to be a significant relationship between the relative **lack of 'mutual dependence'** among Arts and Humanities researchers, their **'task uncertainty'** and the ways in which digital resources are **produced and utilised**. The arrangements for collaborative research and for disseminating research results are **personalised, localised and decentralised**. Informal communication depends on individual groups and specific social networks. Digital resources, where they exist, tend to be field-based and similarly localised. Likewise, there is a corresponding reliance on commercially produced generalist digital resources. We could produce no reliable estimate of what proportion of resources were in proprietary (i.e. commercially-provided, subscription-based or purchased information) as opposed to public-domain (i.e. free to access, generally publicly-funded information) information. Our users were often not aware of the contractual basis on which the information was provided to them. Nor could we estimate how frequently, and for how long, they consulted these resources – the patterns were too varied.

o There is a perception among arts and humanities scholars that within their fields there is little or no collaboration. The reality is substantially different, because while **strong collaborative cultures may not exist**, however, weak ones do and take the form of citations of colleagues' works, routine email correspondence, interaction through conferences and professional society meetings.

o Arts and Humanities **'e-infrastructure'**, apart from the AHDS and RDN subject-portals, tends therefore to be determined at the level of the **employing institution** rather than the field or discipline, or higher.

o We have to take into account a **'counter-culture'** (which we encountered most noticeably in our investigation of some of the disciplines in the Arts, where information resource needs and research agendas are often articulated in terms of 'diametrical difference' to prevailing trends).

o Arts and Humanities researchers want access to information, irrespective of the medium in which it is available. They are used to working in fields where there is a very **mixed economy of resources**, electronic and physical. Journal articles are important, but so are printed books. E-prints (pre- and post-) are markedly less significant than in the physical sciences and engineering. Electronic bibliographical information is therefore of critical importance to Arts and Humanities researchers.

o There is less emphasis on communicating work-in-progress and more emphasis on **formal ways of disseminating information**. There is consequentially less emphasis on lead-times for accessing research-sensitive information and results.

6.1.2 Information about users' awareness and attitudes with respect to currently available online services and tools, including such gateways and portals as current exist.

In general, we encountered a **high and growing level of expectation** as to the availability of materials in digital form. These expectations have been fed by the **exponential growth in the content of Arts and Humanities digital libraries** by the wide variety of different content-creators and contractors.

Generally users were largely **unaware of the possibilities** for data analysis and multimedia data presentation that digitisation offers and were equally **unaware of the extent to which their use of digital resources is tracked and analysed** by content and service providers and employers.

The **internet search engine** emerges from this study as an immensely useful digital resource-discovery tool. Users deployed a variety of proprietary search-engines. Their simplicity and speed appealed to our users, for whom a key determinant in their cost-benefit analysis of resource-discovery tools was whether it saved, rather than cost them time. At the same time, our users were also aware of the limitations of their internet search-engine of choice. Our users told us of their **frustration at its lack of sophistication**. They were **suspicious of its ranking of hits** returned. They were **overwhelmed by the information redundancy** which often accompanies its results. They were, above all, concerned about the fact that search-engines do not search a great deal of digital content that is relevant to their needs; and, equally, they are frustrated by the lack of interoperability between different libraries of digital content.

The issue of **'access'** runs throughout our report. Access to online journals was most often raised; but it frequently occurred also in respect of proprietary digital content of various kinds, specific to particular disciplines. The issue was sometimes presented in terms of a trade-off in resource terms, with our users wanting to see the investment of scarce resources in widening the local access to digital content through licence and content purchase rather than increased investment in resource discovery. At the same time, our research practitioners were aware that 'access' was only fully beneficial when it was linked to enhanced resource discovery, and, in particular, interoperability.

Interoperability was another major theme running through our enquiries. It tended to affect some disciplines more than others. As digital content becomes richer and more diverse, so the independent platforms on which it is consulted multiply. As interoperability becomes more important, so the potential for a next-generation resource-discovery portal grows. While the AHDS and Intute allow their resources to be harvested by other services, **they do not themselves comprehensively harvest available metadata**. For the AHDS this is due to their remit of collecting 'from' not 'for' the research community, while Intute-Arts and Humanities has indicated a general lack of useful metadata available. Intute has RSS news feeds that aggregate news and new collections. End users appear to find this easier to use than Open Archives Initiative (OAI) metadata-harvesting. This is a form of service that is already appreciated by

individual users. This would appear to be a more advantageous route for making data available to commercial harvesters than that provided by the Open Archives Initiative (OAI) metadata-harvesting. The latter has currently received only limited take-up within institutions and none to our knowledge by individuals.

6.1.3 Information about patterns of recent user-activity in relation to the RDN subject hubs and AHDS.

From the wide-range of resource-discovery services and tools used by Arts and Humanities scholars, we investigated user familiarity with and use of these two services in particular. The key feature of the **RDN subject-portals** is their **resource descriptions**. Although our users were clear about the potential importance of **authenticating** digital resources, they were not so sure about the **significance** of the resource descriptions provided by the RDN portals. In particular, they had no sense as to how often they were **up-dated**, the **status** of who had written them, and what **range** of resources they covered. Those that had used the subject portals, took the view that they tended to be useful at the beginning of a research enquiry, but to become rapidly less relevant the more one advanced into a subject. Those that had not used the RDN subject-portals but knew of their existence had evidently formed a view about whether they were **likely** to find anything of relevance to them within it. We conclude from our evidence that the RDN portals are insignificant for most research purposes for the Arts and Humanities practitioner.

AHDS has a similarly low profile among the majority of arts and humanities researchers, although the evidence from AHDS web-logs may well be deceptive. Overall they may under-record some aspects of its usage despite some inflation of usage figures resulting from the inclusion of internal traffic between different servers within the AHDS network as a whole, including network administration calls. Although the number of **resources downloaded** seems to be increasing, none of the participants in our focus groups or questionnaire admitted to having downloaded such collections. Where the AHDS **harvested** data, generally in collaboration with outside partners (as in e.g. Heirport the Historical Environment Information Resources Portal), it plays a significant, perhaps pivotal, role in Arts and Humanities research.

Neither service has a published strategy for consulting users and discovering their needs, although there are examples of good practice in some parts of the AHDS. There are some good collaborative links with other information service-providers in place, but these need to be strengthened. The two services are not currently interacting very well. The RDN subject-portal does not harvest the metadata on AHDS resources comprehensively. While references to each other can be found on their respective sites, neither service promotes the other particularly actively, explains their relationship/differences or provides a quick and easy link to the other.

6.1.4 Information about users' responses to what future portal developments can deliver

Users generally found the current resource-discovery arrangements and services adequate, but were confused about the roles. The evidence is that researchers are more concerned with access to content than functionality.

At the same time, they recognize that the current situation with regard to functionality is not sustainable in the longer term. The importance of interoperability in users' minds was a measure of that realization. The exponential growth in data volume, combined with increasingly complex multilayered information, will make it more necessary to use resources in a complementary way, and simultaneously harder to do so.

Our users responded positively to the possibilities of a **personally-managed** research environment. There were specific, realizable functionalities that they identified as being of direct use to them in carrying forward their research agendas: **workflow management tools** and **resource discovery tools**. Researchers wanted greater personal control over digital resources. They readily perceived the advantages of tools which enabled them to integrate searching the web with searching their own hard-drive. They saw benefits to more developed **bookmarking** features, **personal editing** features, and an **automated copyright management system**. They wanted to be able to **filter** the quality of hit returns, **search distributed databases.** They responded positively to a **web-based news feed feature**, and liked the idea of **RSS feeds** that by-passed personal email accounts.

They were less excited about tools to enable communication and collaboration. The picture that emerged is of researchers who find asynchronous and largely mono-media communication channels such as email, web pages and telephone quite satisfactory. Real-time communications media such as instant relay chat and Grid videoconferencing with integrated computer applications sharing were less appealing. However most respondents declared themselves **happy to collaborate at the basic level of sharing the sources they used.**

Many of the features presented in the demonstrator imply a more sophisticated portal tool than the current gateways provide, and that requires a development in the ICT skills-base of the user-community which it is clearly reluctant to make. The investments made in the ICT skills-base through the Methods Network, ICTguides and training/awareness programmes organised by the AHDS cannot be expected to uplift the skills-base of researchers who do not currently see the need to do so. Whilst this skills-base is likely to improve over time, the potential functionality of portal tools will probably always outstrip it.

6.2 Ways Forward

We see a number of ways forward.

1. An awareness of the **distinctive research culture** with its fears and predilections must be taken into account.

2. The Arts and Humanities research community is not very assertive. Its digital resource-discovery needs have not been very well-voiced. As digital data expands exponentially in our field, and becomes increasingly complex and multi-layered, it is going to become harder to find, and use what we need. The arts and humanities need **strong pan-institutional organisations** that can champion them nationally and internationally. This is a role that AHDS is beginning to play in relation to standards (Brown et al 2006) but it applies also to information resource-discovery needs, including issues of access to content. The AHDS' has a singular focus on arts and humanities. Intute-Arts and Humanities has been established to function as a distinct service for the arts and humanities. The case for a single and coherent resource discovery service for arts and humanities is from the point of view of the user, clear.

3. The increasing provision of **metadata-harvesting** among the information service-providers is an immediate and short-term objective, dominating the agenda of resource-discovery over the next five years. Users are coming to expect much better **linkage** between **online bibliographical resources**, and the **online content** itself. They also want to search across **distributed digital data**. This objective implies:
    o **common metadata standards** [substantially in place]
    o agreed **authentication systems** [emerging, but more work needed]
    o much greater degree of **collaboration** among a wider group of information service-providers than is currently in place (research libraries: archives: museums: government/commercial information-providers, etc) [not in place]
    It is beyond our remit to recommend where such collaboration should come from. But we are convinced that the AHDS has a more important role to play in participating in, and facilitating, such collaborations than it has played in the past.

4. In the medium and longer term (in a five-ten year perspective), it is likely that the **semantic web**, especially when combined with harvesting agents, will provide the easy-to-use tools that many researchers need, at least to some degree. However, for some areas of the Arts and Humanities where "knowledge" is more the result of heuristics and associative thinking, it may be that a more folksonomic approach as exemplified by **Web 2.0** services such as Flickr and steve.museum will be more effective. We are therefore more persuaded in the shorter-term of the possibilities of **Web 2.0** offering a way forward in the form of community-contributed and mediated content. Users do not seem averse to contributing in that way, but the nature of 'mediation' should be recognized. We can see the

possibility of the RDN subject-portals evolving towards **a different mediation role**, with resource-discovery content coming instead from the community itself. In the longer term, there may be a possibility for combining the semantic-web and Web 2.0 approaches, especially if and where discipline-based ontologies emerge as commonly accepted.

5. We can begin to discern the determining characteristics of the resulting information environment as it emerges over the coming decade. It will be:

    o   inclusive
    o   aggregative
    o   personalisable
    o   locally managed
    o   quality-assured
    o   easy to use
    o   community-based
    o   internationally developed

    At various points in this report we have referred to this as a **'managed research environment'**. The use of the term "environment" rather than "portal" is significant here because it does not necessarily entail a single provider. It could comprise a selection of Web portal services, or "portlets", that users draw down to their desk top and configure personally or it may take the form of a pre-configured set embedded within a trusted supplier such as an institutional or professional society web site. Moving towards such an environment should be regarded as a medium-term objective (i.e. three to five years). The current portal providers in the Arts and Humanities do not look like this. But, of course, there are already individual services in the public domain that have some or all of these features and there are recent precedents for the kind of environment we have described. For example, the JISC/LTSN Learning and Teaching Portal Project resulted in a set of web portal services that are embedded in the HE Academy website as a suite of 'Finder' services that could be adopted by other organisations (http://www.heacademy.ac.uk/48.htm).

    We know that Arts and Humanities researchers are prepared to seek out and employ unusual, and 'unauthorised' sources for their information. We also know that they are willing to share useful sources they have discovered themselves. It seems likely that, if researchers come to recognize the existence and utility of such tools and services as these, they will employ them in greater numbers, further undermining the viability of established and 'authorised' services.

6. In the development of such a 'managed research environment' in the Arts and Humanities, there is also scope for collaboration with information system developers, including commercial and international providers. We do not exclude the possibility of UK collaboration in this area with developments currently under Beta-test in 'Google Scholar' to share the costs and manage the delivery. Many

of these tools will need to conform to the international standards that are encouraging British developments to be compatible with a much larger range of applications.

We therefore recommend a scoping study to ascertain the feasibility of such collaboration and the costs of developing a research-directed community-driven subject portal that offers:

- o **Workflow Management tools** that give the researcher greater personal control over digital project resources, especially more evolved **bookmarking features** and some form of automated **copyright management system** to facilitate the growing concern with usage permission and intellectual property rights was also highly valued.
- o **Resource Discovery tools** that provide greater control over web-based resources including the ability to **filter** the quality of hit returns, **search** multiple databases
- o **News feed features** that by-pass personal email accounts, but notify users of conferences, funding, jobs and new research publications.
- o **Collaborative research tools** for social bookmarking, uploading and sharing resources, annotating digital resources, shared document editing, attaching metadata to personally-created digital resources, and contributing to the authentication of digital content.

7. We recommend in the **short term** (one-two years) a much greater collaboration through data-harvesting of the current AHDS and former RDN subject-portals in resource discovery provision and though cross promotion of each others' services.

8. In the **medium term** (three-five years) we recommend that the AHDS and Intute develop a more Web 2.0 compatible profile to enable greater community involvement in resource recommendation, evaluation, creation, selection, sharing and annotation. We also recommend that funding bodies such as JISC and AHRC positively encourage and facilitate the development of interoperable portlets that can be used to embed portal type functionality in institutional and community web sites. An example of this may already be seen in the use of RSS news feeds offered by both services in order to announce news and collections.

9. In the **medium to long term** (five-ten years) we recommend that the AHDS and Intute-Arts and Humanities consider integrating their databases and user interfaces to provide the nucleus of a new, seamless, more comprehensive service in this particular area, one that combines and integrates the core functions of data-archiving, and digital resource harvesting/indexing. This would mean a harmonisation of Web portal services, as opposed to a merging of the two organisations.

List of Tables and Figures

A5	**Work Package 2 Web-Log Analysis Report**	

References Cited and Resources Consulted

(AHDS Archaeology Data Service), Arts and Humanities Data Service (1997), 'AHDS Archaeology
          Annual Report 1996-1997.'
          http://ads.ahds.ac.uk/project/annrpts/1997rpt.html.
--- (1998), 'AHDS Archaeology Annual Report 1997-1998.'
          http://ads.ahds.ac.uk/project/annrpts/1998.html.
--- (1999), 'AHDS Archaeology Annual Report 1998-1999.'
          http://ads.ahds.ac.uk/project/annrpts/1999.html.
--- (2000), 'AHDS Archaeology Annual Report 1999-2000.'
          http://ads.ahds.ac.uk/project/annrpts/2000.html.
--- (2001), 'AHDS Archaeology Annual Report 2000-2001.'
          http://ads.ahds.ac.uk/project/annrpts/2001.html.
--- (2002), 'AHDS Archaeology Annual Report 2001-2002.'
          http://ads.ahds.ac.uk/project/annrpts/2002.html.
--- (2003), 'AHDS Archaeology Annual Report 2002-2003.'
          http://ads.ahds.ac.uk/project/annrpts/2003.html.
--- (2004a), 'AHDS Archaeology Annual Report 2003-2003.'
          http://ads.ahds.ac.uk/project/annrpts/2004.html.
--- (2004b), 'AHDS Archaeology Annual Report 2003-2004.'
          http://ads.ahds.ac.uk/project/annrpts/2004.html.
--- (2005), 'AHDS Archaeology Annual Report 2004-2005.'
          http://ads.ahds.ac.uk/project/annrpts/2005.html.
(AHDS-History), Arts and Humanities Data Service (1997), 'AHDS History Annual Report 1996-1997.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual1996-97.htm.
--- (1998), 'AHDS History Annual Report 1997-1998.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual1997-98.htm.
--- (1999), 'AHDS History Annual Report 1998-1999.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual1998-99.htm.
--- (2000), 'AHDS History Annual Report 1999-2000.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual1999-2000.htm.
--- (2001), 'AHDS History Annual Report 2000-2001.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual2000-01.htm.
--- (2003), 'AHDS History Annual Report 2002-2003.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual2002-03.htm.
--- (2004), 'AHDS History Annual Report 2003-2004.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual2003-04.htm.
--- (2005), 'AHDS History Annual Report 2004-2005.'
          http://ahds.ac.uk/history/about/reports-and-policies/annual/annual2004-05.htm.
(AHDS-Literature Language and Linguistics), 'Arts and Humanities Data Service 'Oxford Text Archive
          Annual Report – 1st August 2000-31st July 2001.'
          http://ahds.ac.uk/litlangling/about/reports/AHDSLLL_Annual%20Report_2000-01.pdf.
--- (2002), 'AHDS Oxford Text Archive Annual Report – 1st August 2001 - 31st July 2002.'
          http://ahds.ac.uk/litlangling/about/reports/AHDSLLL_Annual%20Report_2001-02.pdf.
--- (2003), 'AHDS Literature Language and Linguistics Annual Report - August 2002 – July 2003.'
          http://ahds.ac.uk/litlangling/about/reports/AHDSLLL_Annual%20Report_2002-03.pdf
--- (2004), 'AHDS Literature Language and Linguistics Annual Report - August 2003 – July 2004'
          http://ahds.ac.uk/litlangling/about/reports/AHDSLLL_Annual%20Report_2003-04.pdf.
--- (2005), 'AHDS Literature Language and Linguistics Annual Report August - 2004 – July 2005.'
          http://ahds.ac.uk/litlangling/about/reports/AHDSLLL_Annual%20Report_2004-05.pdf.
(AHDS Performing Arts), 'Arts and Humanities Data Service Performing Arts (2003), Arts and Humanities
          Data Service Performing Arts Annual Report 2003.'
          http://ahds.ac.uk/performingarts/about/reports-and-policies/annual.htm.

(AHDS Visual Arts) (2005a), 'The Digital Picture: a future for digital images in UK arts education (AHDS Visual Arts).'
http://thedigitalpicture.ac.uk/home.html.
(AHDS Visual Arts), Arts and Humanities Data Service (1998a), 'AHDS Visual Arts Annual Report 1997-1998.'
http://ahds.ac.uk/visualarts/about/reports/pr_97-98.htm.
--- (1998b), 'VADS User Needs Survey 1998.'
http://vads.ahds.ac.uk/reports/user_survey/user_survey.html.
--- (1999), 'AHDS Visual Arts Annual Report 1998-1999.'
http://ahds.ac.uk/visualarts/about/reports/pr_98-99.htm.
--- (2000), 'AHDS Visual Arts Annual Report 1999-2000.'
http://ahds.ac.uk/visualarts/about/reports/pr_99-00.htm.
--- (2001), 'AHDS Visual Arts Annual Report 2000-2001.'
http://ahds.ac.uk/visualarts/about/reports/pr_00-01.htm.
--- (2002), 'AHDS Visual Arts Annual Report 2001-2002.'
http://ahds.ac.uk/visualarts/about/reports/pr_01-02.htm.
--- (2003), 'AHDS Visual Arts Annual Report 2002-2003.'
http://ahds.ac.uk/visualarts/about/reports/pr_02-03.htm.
--- (2004), 'AHDS Visual Arts Annual Report 2003-2004.'
http://ahds.ac.uk/visualarts/about/reports/pr_03_04.htm.
--- (2005b), 'AHDS Visual Arts Annual Report 2004-200.'
http://ahds.ac.uk/visualarts/about/reports/Visual%20Arts%20Annual%20Report%202004-2005.pdf.
(AHDS Visual Arts), 'Arts and Humanities Data Service: Visual Arts 'Image Portal Demonstrator'
http://ahds.ac.uk/__print__/visualarts/projects/image-portal/index.htm.
(AHDS) Arts and Humanities Data Service (2003), 'Arts and Humanities Data Service (AHDS) Annual Report 2002-2003.'
http://www.ahds.ac.uk/history/about/reports-and-policies/index.htm.
--- (2003), 'Image Portal Demonstrator.'
http://ahds.ac.uk/visualarts/projects/image-portal/index.htm.
--- (2004a), 'The Arts and Humanities Data Service Strategic Plan 2002-2005'(2004).
http://ahds.ac.uk/about/reports-and-policies/index.htm. http://ahds.ac.uk/documents/exec-strategy-02to05-v1.pdf.
--- (2004b), 'Arts and Humanities Data Service (AHDS) Annual Report 2003-2004'
http://www.ahds.ac.uk/history/about/reports-and-policies/ahds-annual-report-2003-04.pdf.
--- (2004c), 'AHDS Strategic Plan 2005-2007.'
http://ahds.ac.uk/documents/exec-strategy-05to07.pdf
--- (2005), 'Arts and Humanities Data Service (AHDS) Annual Report 2004-2005.'
http://www.ahds.ac.uk/history/about/reports-and-policies/ahds-annual-report-2004-05.pdf.
(AHRC), 'Arts and Humanities Research Council 'The Bright Path: Strategy for the Arts and Humanities Research in the UK.'
http://www.ahrc.ac.uk/ahrb/website/news/publications.asp.
http://www.ahrb.ac.uk/ahrb/website/images/4_94380.pdf.
--- (2005), 'Annual Report 2004-05.'
http://www.ahrc.ac.uk/ahrb/website/images/4_96568.pdf.
--- (2006), 'ICT in the Arts and Humanities.'
www.ahrcict.rdg.ac.uk/index.htm.
(ANIR) Working Group on Access to Networked Information Resources (1994), 'Report of the Working Group on Access to Networked Information Resources', *Journal of Information Networking,* 2 (3).
(HERA) Humanities in the European Research Area 'Homepage of the Humanities in the European Research Area (HERA).'
http://www.heranet.info/Default.aspx?ID=102
http://www.nwo.nl/nwohome.nsf/pages/NWOP_6D4C8L_Eng.
(HESA) Higher Education Statistics Agency 'Staff Tables 2003-2004'
http://www.hesa.ac.uk/holisdocs/pubinfo/staff.htm.
--- (2004b), 'Student Tables 2003-2004.'

http://www.hesa.ac.uk/holisdocs/pubinfo/stud.htm.
--- 'HESA On-line Information Service.'
        http://www.hesa.ac.uk/holisdocs/home.htm.
(IATH), Institute for Advanced Technology in the Humanities (2005), 'Summit on Digital Tools for the
        Humanities', paper given at Summit on Digital Tools for the Humanities. Charlottesville,
        Virginia.
        http://www.iath.virginia.edu/dtsummit/SummitText.pdf.
(JISC) Joint Information Systems Committee (1999a), 'Adding value to the UK's learning, teaching and
        research resources: the Distributed National Electronic Resource (DNER).
        http://www.computing.dundee.ac.uk/projects/dmag/bobby/jisc/194b.html
--- (1999b), 'Description of the DNER'   http://www.jisc.ac.uk/index.cfm?name=dner_description.
--- (2002a), Adding Value to the UK's Learning, Teaching and Research Resources: the Distributed
        National Electronic Resource (DNER).'
        http://www.jisc.ac.uk/index.cfm?name=dner_adding_value.
--- (2002b), 'JISC Five Year Strategy, 2001-2005'
        http://www.jisc.ac.uk/index.cfm?name=strategy0105_summary.
--- (2003a) 'A Brief History of JISC.'
        http://www.jisc.ac.uk/index.cfm?name=about_history.
--- (2003b), 'Description of the Distributed National Electronic Resource (DNER).'
        http://www.jisc.ac.uk/index.cfm?name=dner_description.
--- (2003c)'JISC Information Environment: Portals.'
        www.portal.ac.uk.
--- (2003d), 'JISC Information Environment Portals: Investigations into User Requirements and
        Sustainability: INVITATION TO TENDER.'
        http://www.jisc.ac.uk/uploaded_documents/PortalsITTOct03.doc
        http://www.jisc.ac.uk/index.cfm?name=funding_portals.
--- (2003e), 'Portals: Frequently Asked Questions.'
        http://www.jisc.ac.uk/index.cfm?name=ie_portalsfaq.
 --- (2006a), 'Automating the Selection of Resources for eLearners', Cross-Institutional Use of E-Learning
        to Support Lifelong Learners Programme
        http://www.jisc.ac.uk/index.cfm?name=funding_circular03_06.
--- (2006b), 'Invitation to Tender: The Arts and Humanities Data Service (AHDS) Review and User Survey'
        http://www.jisc.ac.uk/index.cfm?name=funding_ahds_review_and_user_survey
        http://www.jisc.ac.uk/uploaded_documents/ITT%20AHDS%20Review%20and%20User%20Surv
                ey%20April%202006.DOC.
--- (2006c), 'Notes from a Consultation Workshop on the Virtual Research Environments Programme',
        paper given at London, February 2006.
        http://www.jisc.ac.uk/uploaded_documents/VRE_Workshop_28Nov_notes.doc.
-- (2006d),'Virtual Research Environments Programme.'
        http://www.jisc.ac.uk/index.cfm?name=programme_vre.
(JISC MU) Joint Information Services Committee Monitoring Unit (2005), 'Workflow Solution to the
        Preservation of Digital Media', 'Survey Report: Content Services 2004-05.'
        http://www.mu.jisc.ac.uk/reports/surveys/.
        http://www.mu.jisc.ac.uk/reports/surveys/ncs2004/ncs2004-05-survey-report.pdf.
(MIMETIC) Meeting Memory Technology Informing Collaboration 'Project Homepage.'
        http://www.memetic-vre.net/.
(SOSIG) Social Science Information Gateway (2002), 'Evaluation and Feedback from User Communities.'
        Personal Communication.
Aarseth, E. (2003), 'The Field of Humanistic Informatics and its Relation to the Humanities.'
        http://www.hf.uib.no/hi/espen/HI.html.
Abbott, D. and Beer, E. (2006), 'Getting to Know Our Audience. AHDS Performing Arts Scoping Study.'
        http://www.hatii.arts.gla.ac.uk/ahds-pa/.
Accenture (2003), 'University Portals: Opening Doors to Better Service.'
        http://www.gre.ac.uk/ils/cis/portalproject/universityportals.pdf.
Adler, M., and Ziglio, E. (1996), 'Gazing into the oracle' (Bristol, PA: Jessica Kingsley).

Agarwal, S. and Browning, P. (2004), 'Focusing on users: case study of portal user requirements analysis and user-testing sessions at the University of Bristol' *Online Information* 2004 Proceedings. http://www.online-information.co.uk/2004proceedings/wedpm/agarwal_browning.pdf.

Allan, R., Awre, C. Baker, M., Fish, A (2003), 'Portals and Portlets 2003', *Technical Report UKeS-2004-06.* http://www.nesc.ac.uk/technical_papers/UKeS-2004-06.pdf.

American Council of Learned Societies (1997), 'The Transformation of Humanistic Studies in the Twenty-first Century: Opportunities and Perils', Occasional Paper No. 40. http://www.acls.org/op40.htm.

--- (2005), 'Our Cultural Commonwealth.' http://www.acls.org/cyberinfrastructure/cyber_report.htm.

Anderson, I.G. (2004), 'Are you Being Served? Historians and the Search for Primary Sources.' *Archivaria* 58 (Fall). http://archivists.ca/downloads/documentloader.aspx?id=3661.

Anderson, S. (2004), E-Science (E-Research Expert Seminar: Report on Proceedings. Senate House London. www.ahds.ac.uk. http://www.ahds.ac.uk/e-science/e-science-seminar-2004.pdf#search=%22%20%22E-Science%20(E-Research%20%20Expert%20Seminar%3A%20Report%20on%20Proceedings%22%22.

Anderson, S., Dunn, S., Hughes, L. (2005), 'VREs In the Arts and Humanities', paper given at All Hands Meeting. http://www.ahrcict.rdg.ac.uk/info/vre/AHM05_paper.pdf.

Armstrong, CJ; Lonsdale, R., Stoker, D., and Urquhart, C. (2000), JUSTEIS Project as detailed below:

--- (2000), 'JISC Usage Surveys: Trends in Electronic Information Services.'

--- (2000) 'Strand A: A general survey of end users of all electronic information services.'

--- (2000), 'Strand C: A general survey of electronic information services provision.'

--- (2000), 'Final Report - 1999/2000 Cycle.' http://www.dil.aber.ac.uk/dils/research/justeis/cyc1rep0.htm.

Asensio, M. (2003), 'Final Study Report JISC User Requirement Study for a Moving Pictures and Sound Portal.' http://www.jisc.ac.uk/uploaded_documents/MPSportaluserreqs.doc.

Austin, T., Pinto, F., Richards, J. and Ryan, N. (2001), 'Joined up writing: an internet portal for research into the historic environment', paper given at Forthcoming paper in CAA 2001: Proceedings of Computer Applications and Quantitative Methods in Archaeology Conference., Gotland University, Visby, Gotland - Sweden, April 2001. http://www.cs.kent.ac.uk/pubs/2001/1261/ http://www.cs.kent.ac.uk/pubs/2001/1261/content.pdf.

Awre, C., Hanganu, G., Ingram, C., Brett, T., and Dolphin, I. (2005), 'The CREE Project: Investigating User Requirements for Searching within Institutional Environments ', *D-Lib Magazine,* 11 (10)

Barab, S. A., and Krishner, D. (2001), 'Guest editors' introduction: 'Rethinking methodology in the learning sciences', *Journal of the Learning Sciences,* 10 (1&2), 5-15.

Becher, T. (1987), 'Disciplinary Discourse', S*tudies in Higher Education,* 12 (3), 261-74. http://taylorandfrancis.metapress.com/(hgezqt55un1kpa553ajmx045)/app/home/contribution.asp?referrer=parent&backto=issue,3,10;journal,67,91;linkingpublicationresults,1:104673,1. http://taylorandfrancis.metapress.com/media/p3dgtpwwwg62076klqfw/contributions/k/0/k/l/k0kl85421441277x.pdf.

--- (1994), 'The Significance of Disciplinary Differences', *Studies in Higher Education,* 19 (2), 151-61. http://taylorandfrancis.metapress.com/(ukbale45kiwotsyajxedb2vb)/app/home/contribution.asp?referrer=parent&backto=issue,4,12;journal,47,91;linkingpublicationresults,1:104673,1. http://taylorandfrancis.metapress.com/media/h83augyqqjdufcxwhkdr/contributions/n/k/0/0/nk00303x8328m446.pdf.

Becher, T. and Trowler, P. (2001), *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines* 2nd ed., Buckingham: Open University. http://mcgraw-hill.co.uk/openup/chapters/0335206271.pdf.

Beckels, B., Brostoff, S., and Ballard, S. (2004), 'A First Attempt: Initial Steps Toward Determining Scientific Users' Requirements and Appropriate Security Paradigms for Computational Grids.' Proceedings of the Workshop on Requirements Capture for Collaboration in e-Science. Edinburgh, UK.
http://www.escience.cam.ac.uk/papers/req_analysis/first_attempt.pdf.

Beckles, B. (2004), 'User requirements for UK e-Science grid environments.'
http://www.allhands.org.uk/2004/proceedings/papers/251.pdf.

Berman F., Fox G.C., Hey A.J.G. (eds.) (2003a), *Grid Computing: Making the Global Infrastructure a Reality*. John Wiley & Sons.

Berman F., Fox G.C., Hey A.J.G. (2003b), 'The Grid: past, present, future' in Berman et al 2003a pp 9-50

Berners-Lee, T., Hendler, J. and Lassila, O. (2001), 'The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities'. In *Scientific American*, 17 May 2001.
http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Beyer, H. and Holtzblatt, K (1998), *Contextual Design: Defining Customer-Centered Systems* London: Morgan Kaufmann.

Biblio-Tech Review 'Z39.50 Overview.'
http://www.biblio-tech.com/html/z39_50.html.

Bowman, A., Crowther, C., Kirkham, R., and Pybus, J. (2005), 'Building a Virtual Research Environment for the Humanities: PROJECT PLAN.'
http://bvreh.humanities.ox.ac.uk/BVREH-ProjectPlan-web.pdf.

Brewer, J. and Kilbride, W.G. (2005), 'HEIRNET User Survey 2005 Report and analysis (Council of British Archaeology).'

British Library (1993), 'Information Technology in Humanities Scholarship' British Library R & D Report 6097.

Brockman, W.S., Neumann, L., Palmer, C.L. and Tidline, T. (2001), 'Scholarly Work in the Humanities and the Evolving Information Environment.'
http://www.clir.org/
http://www.clir.org/pubs/reports/pub104/pub104.pdf.

Brown, A. L. (1992), 'Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings', *The Journal of the Learning Sciences,* 2, 141-78.

Brown, S and Liang, V. (2005), 'Work Package 3 User Needs: Report'.
http://aria.dmu.ac.uk/documents/WorkPackage3.pdf.

Brown, S., Bryson, J., Greengrass, M. and Ross, R. (2006), 'Arts and Humanities Data Service (AHDS) Review and Final Report', unpublished report to Joint Information Systems Committee and Arts and Humanities Research Council.

Burnard, L.and Short, H. (1994), 'An Arts and Humanities Data Service. Report of a Feasibility Study' Commissioned by the Information Services Sub-Committee of the Joint Information Systems Committee of the Higher Education Funding Councils (Oxford).

Burnard, L. (1999), 'Is Humanities Computing an Academic Discipline? or Why Humanities Computing Matters.'
http://www.iath.virginia.edu/hcs/burnard.html.

Butters, G. (2003), 'What Features in a Portal?' *Ariadne* (35).
http://www.ariadne.ac.uk/issue35/butters/.

Currier, S. and Campbell, L. (2002), 'Evaluating Learning Resources for Reusability: The DNER & Learning Objects Study', Australasian Society for Computers in Learning in Tertiary Education conference proceedings. Aukland 2002.
http://ascilite.org.au/conferences/auckland02/proceedings/papers/059.pdf

Case, D. (2002), 'Looking for Information: A Survey of Research on Information Seeking Needs and Behaviour', ed. Harold Borko (Library and Information Science; London: Academic Press).

Chae, B. and Poole, M (2005), 'Mandates and Technology Acceptance: A Tale of Two Enterprise Technologies', *Journal of Strategic Information Systems,* 14, 147-66.
http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6VG3-4G54HJ6-1-1&_cdi=6027&_user=128590&_orig=search&_coverDate=06%2F30%2F2005&_qd=1&_sk=999859997&view=c&_alid=442604047&_rdoc=1&wchp=dGLbVlz-zSkWb&md5=16225d99f0402dca64c9906fcd303ce2&ie=/sdarticle.pdf.

Chambers, S. and Tiedau, U. (2002), 'Research goes virtual', *Library and Information UPDATE.*
http://www.cilip.org.uk/publications/updatemagazine/archive/archive2002/july/update0207c.htm.

Chappell, D.A and Jewell, T. (2002), *Java Web Services*. O'Reilly and Associates.

Christiansen, L.(2004), 'Blueprint for the European Research Observatory for the Humanities and Social Sciences (EROHS).'
http://www.heranet.info/Admin/Public/DWSDownload.aspx?File=Files%2fFiler%2fERCH%2fRISSH_-_ESFRIJUNE2004-version.doc.

Clark, J. (2003), 'Overview and Requirements Analysis: General Portals Functionality for SAD I', Subject Portals Project Phase I Documents.'
http://www.portal.ac.uk/spp/documents/phase1/projectplan/spdpoverview.doc
http://www.portal.ac.uk/spp/documents/phase1/.

--- (2001b) 'Subject Portals', *Ariadne,* (29)
http://www.ariadne.ac.uk/issue29/clark/intro.html

Cliffe, P. and Powell , A. (no date) 'UK Z39.50 Directory'
http://www.ukoln.ac.uk/distributed-systems/zdir/.

Cloiser, A. (2003), 'Interim User Requirements Report for the JISC Information Environment Service Registry.'
http://iesr.ac.uk/inerimreqs.html.

Collins, A. (ed.), (1992), *Toward a design science of education*. E. Scanlon & T.

Committee on a KNAW Research Institute for e-Science, (2003), 'Building the KNAW International Research Institute on e-Science Studies in the Humanities and Social Sciences (IRISS).'
http://www.knaw.nl.
www.knaw.nl/publicaties/pdf/90000111.pdf

Condron, F., Richards, J., Robinson, D., Wise, A. (1999), 'Strategies for Digital Data - A Survey of User Needs.'
http://www.eng-h.gov.uk/archcom/projects/.
http://www.eng-h.gov.uk/archcom/projects/summarys/html98_9/2178main.pdf.

Cooper, Ian (2004), 'Survey of the JISC Network Content Services 2004/05.'
http://www.mu.jisc.ac.uk/reports/surveys/ncs2004/.

Cox, A. (2004), 'Building Collaborative eResearch Environments.'
http://www.jisc.ac.uk/index.cfm?name=event_report_eresearch.

--- (ed.), (2006), *Portals: People, Processes and Technology*. Oxford: Facet.

Creswell, J. (2002), *Research Design: Qualitative, Quantitative and Mixed Method Approaches*. London: Sage.

Crouchley, R and Fish, A. (2004a), 'Roadmap for a UK Virtual Research Environment: Report of the JCSR VRE Working Group.'
http://tyne.dl.ac.uk/Sakai/sakai_doc/node11.html.
http://www.jisc.ac.uk/uploaded_documents/VRE%20roadmap%20v4.pdf.

--- (2004b), 'SAKAI Evaluation Exercise: A Report to JISC--Role of Portals in a Virtual Research Environment.'
http://tyne.dl.ac.uk/Sakai/sakai_doc/node11.html.

Cultural Heritage Consortium (2002), 'Heirnet: Historic Environment Information Resources Network. Users and their Uses of Heirs.'
http://www.britarch.ac.uk/HEIRNET/publications.html.
http://www.britarch.ac.uk/HEIRNET/users.PDF.

De Rosa, C. Dempsey, L. and Wilson, A. (2004), '2003 Environmental Scan: Research and Learning Landscape.'
http://www.oclc.org/reports/escan/toc.htm.
http://www.oclc.org/reports/escan/downloads/research.pdf.

Debus, M. (1995), *Methodological Review: A Handbook for Excellence in Focus Group Research* Washington, DC: Academy for Educational Development.

Denzin, N.K., and Lincoln, Y.S. (ed.), (1994), *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage.

De Roure, D., Baker, M.A., Jennings, N. R. and Shadbolt N.R. (2003a), 'The evolution of the Grid.' in Berman, et al 2003a 65 – 100.

De Roure D., Jennings N. R., Shadbolt N.R. (2003b), The Semantic Grid: a future e-Science infrastructure.

In Berman et al 2003a pp 438 – 470.

Design-Based Research Collective (2003), 'Design-Based Research: An Emerging Paradigm for Educational Inquiry'. *Educational Researcher*, Vol. 32.1, pp. 5-8.

Dimitrova, Maia 'Virtual Research Environments Programme.' http://www.jisc.ac.uk/index.cfm?name=programme_vre.

Dolphin, I.; Miller, P. and Sherratt, R. (2002), 'Portals, PORTALs Everywhere', *Ariadne,* (33). http://www.ariadne.ac.uk/issue33/portals/.

Duffy, C. and Kemp, S. (2004), 'What are the ICT needs of practice-based researchers in the visual and performing arts?' paper given at AHRC ICT Expert Seminars, London. http://www.ahrcict.rdg.ac.uk/activities/expert_seminars/. http://www.ahrcict.rdg.ac.uk/activities/expert_seminars/creative_and_performing_arts.pdf.

Dunn, S. (2005), 'Gated gardens? Virtual Research Environments and cross-disciplinary challenges for e-Research', paper given at Digital Resources for the Humanities 2005, Lancaster University. http://ahds.ac.uk/drh2005/viewabstract.php?id=46&cf=1.

Dunn, S. and Dunning, A. (2005b), 'AHRC Research Centres and the Use of ICT.' http://www.ahrcict.rdg.ac.uk/info/centres_projects/phase1.pdf.

Dunning, A. (2004), 'The AHDS is Evolving: Changes at the Arts and Humanities Data Service', *Ariadne,* 38. http://www.ariadne.ac.uk/issue38/ahds/.

ESRC (2005), 'Review of International Data Resources and Needs ' http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/Commissioning_updates/index 86.aspx. http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/IRDRN_Brief_tcm6-8040.pdf.

Fernie, K. (2003a), 'Revealing collections: discovery, access and interoperability.'

--- (2003b), 'SMRs, Users and the Web and Users and their Uses of HEIRs', *Historic Environment Record News,* 1.

Ferry, A. (1997), '1996 Survey of User Information Needs and Search Methods Results: Art, Design, Architecture & Media Information Gateway.' http://adam.ac.uk/adam/reports/survey/.

Finke, A. (1995), *How to sample in surveys.* London: Sage.

Follett, B. (1993), 'Joint Funding Councils' Libraries Review Group (The Follett Report).' http://www.ukoln.ac.uk/services/papers/follett/report/

Foster I., Kesselman C. and Tuecke S. (2001), 'The anatomy of the Grid.' *International Journal of High Performance Computing Applications* Sage– reprinted in Berman et al 2003a 171 – 197,

Foster I., Kesselman C., Nick J.M.and Tuecke S. (2002), 'The physiology of the Grid.' The Global Grid Forum – reprinted in Berman et al 2003a 217 – 249.

Franklin, T. (2004), 'Portals in Higher Education: concepts and models.' http://www.obhe.ac.uk/products/reports/publicaccesspdf/February2004.pdf.

Fraser, M. (2005), 'Virtual Research Environments: Overview and Activity', *Ariadne,* (44). http://www.ariadne.ac.uk/issue44/fraser/.

--- (2006), '20.112 Humbul: The final regeneration.' Humanist Discussion Group, Vol. 20, No. 112. http://www3.iath.virginia.edu/lists_archive/Humanist/v20/0111.html

Frechtling, J. and Sharp Westat, L. (1997), 'User-Friendly Handbook for Mixed Method Evaluations' http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/START.HTM

Fry, Jenny (2004), 'The Cultural Shaping of ICTs within Academic Fields: Corpus-based Linguistics as a Case Study', *Literary and Linguistic Computing*, 3. http://www.swetswise.com/swetsfo/swproxy?url=http%3A%2F%2Fwww.oxfordjournals.org%2F ViewPDF%2Flitlin%2Fhdb%2FVolume_19%2FIssue_03%2Fabstracts%2F190303.sgm &ts=1141311512924&cs=3536757245.

--- (2006), 'Scholarly Research and Information Practices: A Domain Analytic Approach', *Information Processing and Management,* 42 (1), 299-316. http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6VC8-4DVVXJY-1- 1&_cdi=5948&_user=128590&_orig=search&_coverDate=01%2F31%2F2006&_qd=1& _sk=999579998&view=c&_alid=443570996&_rdoc=1&wchp=dGLbVzb- zSkWW&md5=505f16844e9e1106d89c6056e2fd6637&ie=/sdarticle.pdf.

Gannon, D. ;G. Fox, M. Pierce, B. Plale, G. von Laszewski, RG C. Severance, J. Hardin, J. Alameda, M. Thomas, J. Boisseau (2003), 'Grid Portals: A Scientist's Access Point for Grid Services (DRAFT 1).'
http://forge.gridforum.org/projects/ggf-editor/document/GCE-Portal-working-draft/en/1/GCE-Portal-working-draft.pdf>.

Gardener, M. (2001), 'Portals--their role in the emerging networked economy.'
http://www.essex.ac.uk/chimera/content/Pubs/pubs/EURESCOM-Gardner-portals.pdf.

Gerrard, D. (2005), 'Investigating the Application of a Grid Computing "Workflow" to standalone photo-historical data sources', (DeMontfort University).

Goldfisher, K (1993), 'Modified Delphi: A Concept for Product Forecasting', *Journal of Business Forecasting,* Winter.

Goodman, L. and Milton, K. (ed.), (2004), 'A Guide to Good Practice in Collaborative Working Methods and New Media Tools Creation' Catherine Owen (AHDS Guides To Good Practice).
http://www.ahds.ac.uk/creating/guides/new-media-tools/index.htm.

Goodwin, C., Schwartz, M., Nielsen, J. (2005), 'Usability of Intranet Portals: A Report from the Trenches: Experiences from Real-Life Portal Projects. Usability of Intranet Portals: Executive Summary.'
http://www.nngroup.com/reports/intranet/portals/summary.html.

Grout and Rymer, (1998), 'VADS User Needs Survey 1998: Report.'
http://vads.ahds.ac.uk/reports/user_survey/user_survey.html

Greenbaum, T.L. (1993), *The Handbook of Focus Group Research.* New York: Lexington Books.

Greenstein, D. (1998a), 'The Arts and Humanities Data Service Three Years' On', *D-Lib Magazine*, 12.
http://www.dlib.org/dlib/december98/greenstein/12greenstein.html.

--- (1998b), 'The Arts and Humanities Data Service Three Years' On', *D-Lib Magazine*.
http://www.dlib.org/dlib/december98/greenstein/.

Greenstein, D. and Trant J. (1996), 'AHDS: Arts and Humanities Data Service', *Ariadne,* (4).
http://www.ariadne.ac.uk/issue4/ahds/.

Guy, M. (2003) 'User Testing Report**.'** Subject Access to the DNER (SAD I).'
http://www.portal.ac.uk/spp/documents/testing/phase1/usertestingreportv3.doc.

Harely, Diane, Jonathan Henke, Ian Miller, Alison Head, David Nasatir, Jing Guo and Xi Sheng (2004), 'The Use of Digital Resources in Humanities and Social Science Undergraduate education.'
http://digitalresourcestudy.berkely.edu/.
http://digitalresourcestudy.berkeley.edu/pdf/digitalresourcestudy_final_report_exec_summ.pdf.

Harley, D. (2006), 'Use and Users of Digital Resources: A Focus on Undergraduate Education in the Humanities and Social Sciences.'
http://cshe.berkeley.edu/research/digitalresourcestudy/report/digitalresourcestudy_final_report.pdf

Helmer, O. (1977), 'Problems in futures research: Delphi and causal cross-impact analysis', *Futures*, 17-31.

Hepworth, M (1998), 'Investigating methods for understanding user requirements for information products', *Information Research,* 4 (2).
http://www.informationr.net/ir/4-2/isic/hepworth.html.

Howcroft, D., Newell, S. and Wagner, E. (2004), 'Understanding the Contextual Influences on Enterprise System Design, Implementation, Use and Evaluation', *Journal of Strategic Information Systems* 13, 271-77.

Katz, S. (2003), 'Why Technology Matters: the Humanities in the 21st Century', paper given at the Wisbey Lectures, King's College, University of London, 16 October 2003.
http://www.princeton.edu/~artspol/occaspap/Occasional%20Paper-Katz-10-16-03.pdf.

Kirkham, R.and Pybus, J. (2005), 'Building a Virtual Research Environment for the Humanities Interim Results of the User Survey.'
http://bvreh.humanities.ox.ac.uk/BVREH_Interim_Results_of_User_Survey_Report_web.pdf.

Kling, R., Spector, L. and McKim, G. (2002), 'Locally Controlled Scholarly Publishing Via the Internet: The Guild Model', *Journal of Electronic Publishing,* 8 (1).
http://www.press.umich.edu/jep/08-01/kling.html.

Klyne G., Carroll J.J., McBride B. (eds) (2004), 'Resource Description Framework (RDF): Concepts and Abstract Syntax.' *W3C.*
http://www.w3.org/TR/rdf-concepts/

Krueger, R.A. (1988), *Focus Groups: A Practical Guide for Applied Research.* Newbury Park, CA: Sage.

Lamb, A. (2004) 'Web Portals: Rabbit Holes to Grand Gateways.'

http://www.eduscapes.com/sessions/portals/.
Li, S. and Wood, W. (2005), 'Use of Portals in the Academic World.'
        http://www.educause.edu
        http://www.educause.edu/ir/library/pdf/CSD3738.pdf.
Lindstone, H., and Turoff, M (ed.), (1975), *The Delphi Method.* Addison Wesley.
Lonsdale, R E, D A Stoker and C J Urquhart (2001), 'JISC Usage Surveys: Trends in Electronic
        Information Services.'
        http://www.dil.aber.ac.uk/dils/research/justeis/jisctop.htm.
McCarty, W. (1989), 'Humbul.' Humanist Mailing List, Vol. 2, No. 552.
        http://lists.village.virginia.edu/lists_archive/Humanist/v02/0077.html.
McCarty, W. and Short, H (2002), 'A Roadmap for Humanities Computing.'
        http://www.allc.org/reports/map/.
McCarty, W. and Kirschenbaum,,M (2003), 'Institutional models for humanities computing.'
        http://www.allc.org/imhc/.
McCarty, W. (1999), 'Humanities Computing as Inter-discipline', paper given at Is Humanities Computing
        an Academic Discipline?' An Interdisciplinary Seminar, Institute for Advanced Technology in the
        Humanities Charlottesville, Virginia, USA.
        http://www.kcl.ac.uk/humanities/cch/wlm/essays/inter/.
McLeod, M. (2004), 'User-Centered Product Creation in Electronic Publishing: Good Practice Models ', in
        Lizbeth Goodman and Katherine Milton (ed.), A Guide to Good Practice in Collaborative
        Working Methods and New Media Tools Creation (Glasgow: AHDS Performing Arts).
Metcalfe, R. and Manning, P. 'The Requirements for User Profiling to support Portal functionality.'
        http://www.portal.ac.uk/spp/documents/testing/phase1/userscenarios/profiling.doc.
Miller, P. and Greenstein, D. (1997), 'Discovering Online Resources Across the Humanities: A Practical
        Implementation of the Dublin Core.'
        http://ahds.ac.uk/public/metadata/discovery.html.
Mitchell, V (1992), 'Using Delphi to Forecast New Technology Industries', *Marketing Intelligence and
        Planning,* 10 (2).
Moffat, M. (2003), 'Summary of Portal Features Survey and Portal Consultancy Groups.'
        http://www.eevl.ac.uk/public/ASP/info/.
--- (2003b), 'Humbul user testing – tested by 3 undergraduates; 3 postgraduates; three library staff; and 1
        lecturer.'
        http://www.portal.ac.uk/spp/documents/testing/phase1/user/Collated_results_of_Humbul_user_tes
                ting.doc
        http://www.portal.ac.uk/spp/documents/testing/phase2/SPP_Testing_Round3_V1.doc.
Morgan, D.L. (ed.), (1993), *Successful Focus Groups: Advancing the State of the Art.* Newbury Park,
        California: Sage.
Network, Research Information 'Discovery Services: User behaviour, Perceptions and Needs.'
        http://www.rin.ac.uk/?q=user-behaviour-perceptions-and-needs.
Nicholas, D., Huntington, P. & Williams, P. (2002), 'Evaluating metrics for comparing the use of web sites:
        a case study for two consumer health web sites', *Journal of Information Science,* 28, 63-75.
Nicholas, D., Huntington, P. & Williams, P. (2004), 'Digital consumer health information and advisory
        services in the UK: a user-evaluation and sourcebook'
        http://ciber.soi.city.ac.uk/dhrgreports.php.
Nicholas, D., Huntington, P. and Watkinson, A. (2003a), 'Digital journals, big deals and online searching
        behaviour: a pilot study', *Aslib Proceedings,* 55, 84-109.
Nicholas, D., Huntington, P., Lievesley, N. & Wasti, A. (2000), 'Evaluating consumer Web site logs: case
        study The Times/Sunday Times Web site', *Journal of Information Science,* 26, 399-411.
Nicholas, D., Huntington, P., Rowlands, I., Russell, B. & Cousins, J. (2003b), 'Opening the digital box:
        what deep log analysis can tell us about our digital journal users.' Conference in Charleston, SC.
Nicholas, D., Huntington, P., Williams, P. & Dobrowolski, T. (2004b), 'Re-appraising information seeking
        behaviour in a digital environment: bouncers, checkers, returnees and the like', *Journal of
        Documentation,* 60, 24-39.
Notay, B. (no date), 'JISC Presentation Programme.'
        http://www.jisc.ac.uk/index.cfm?name=programme_presentation.

Palmer, C.and Neumann, L. (2002), 'The Information Work of Interdisciplinary Humanities Scholars:
        Exploration and Translation', *Library Quarterly,* 72 (1), 85-117.
Pearce, L. (2003a), 'Apart from the weather, I think it's a good idea: Stakeholder Requirements for
        Institutional Portals', *Ariadne* (35).
        http://www.ariadne.ac.uk/issue35/pearce/.
--- (2003b), 'Defining users and their needs: the PORTAL project work in progress', *SCONUL Newsletter*
        Summer/Autumn.
        http://www.sconul.ac.uk/pubs_stats/newsletter/29/6.RTF.
--- (2003c), 'Institutional Portals: A Review of Outputs.'
        http://www.fair-portal.hull.ac.uk/downloads/iportaloutputs.pdf.
Pearce, L., Carpenter, L., Martin, R. (2003d), 'Stakeholder Requirements for Institutional Portals.'
        http://www.fair-portal.hull.ac.uk/
        http://www.fair-portal.hull.ac.uk/downloads/stakereq.pdf.
Peterson, E. and York, V. (2003), 'User-Evaluation of the Montana Natural Resource Information System
        (NRIS)', *D-Lib Magazine* 7/8.
        http://www.dlib.org/dlib/luly03/peterson/07peterson.html.
Phelps, T. and Watry, P.B. (2005), 'A No-Compromises Architecture for Digital Document Preservation.'
        9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL
        2005) Vienna, Austria.
        http://multivalent.sourceforge.net/Research/Live.pdf#search=%22%20%22a%20no%20compromi
                ses%20architecture%20for%20digital%20document%20preservation%22%22.
Pinfield, S. and Dempsey, L. (2001), 'The Distributed National Electronic Resource (DNER) and the hybrid
        library', *Ariadne,* 26.
        http://www.ariadne.ac.uk/issue26/dner/.
Pinto, F. and Fraser, M. (2003), 'The SPP and its Functionality ', paper given at Investigations into Portals
        seminar, Oxford University Computing Services' May 2003.
        http://www.portal.ac.uk/spp/documents/phase1/presentations/oucs/TheSPP_functionality_files/fra
                me.htm.
Powell, A (2001), 'RSLP (Research Support Libraries Programme) Collection Description.'
        http://www.ukoln.ac.uk/metadata/rslp/.
--- (2006), 'JISC Information Environment Architecture.'
        http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/.
Powers, S. (2003), *Practical RDF*. O'Reilly and Associates.
Rogers, E. (1995), *Diffusion of Innovations.* 4th ed., London: Free Press.
Rosensweig, R. (2006), 'Can History Be Open Source? Wikipedia and the Future of the Past', *The Journal
        of American History,* 93 (1), 117-46.
        http://chnm.gmu.edu/resources/essays/d/42.
Ross, S. (2004), 'The Role of ERPANET in Supporting Digital Curation and Preservation in Europe', *D-Lib
        Magazine*, 7/8.
        dlib.org/dlib/july04/ross/07ross.html.
Ross, S., Anderson, I., Green, D., Albrecht, K. (2003), 'The NINCH Guide to Good Practice in the Digital
        Representation and Management of Cultural Heritage Materials.'
        http://www.nyu.edu/its/humanities/ninchguide/XII/
        http://www.ninch.org/guide.pdf#search=%22%20%22the%20NINCH%20Guide%20to%20Good
        %20Practice%20in%20the%20Digital%20Representation%20and%20Management%20of%20Cul
        tural%20Heritage%20Materials%22%22.
Rowley, J. (2001), 'JISC User Behaviour Monitoring and Evaluation Framework', *Ariadne*, 30.
        www.ariadne.ac.uk/issue30/jisc.
Sandoval, W. A., and Bell, P. (2004), 'Design-based research methods for studying learning in context:
        Introduction', *Educational Psychologist,* 39, 199-201.
Schopf, J. and Newhouse, S. (2005), 'Real Users –Real Requirements: 25 Conversations with UK eScience
        Projects', paper given at UK eScience All Hands Meeting, Nottingham, September.
        http://www-unix.mcs.anl.gov/~schopf/Talks/25proj-ahm-sept05.pdf.
Schreibman, S., Siemens, R. and Unsworth, J (ed.), (2004), *A Companion to Digital Humanities.* Oxford:
        Blackwell Publishing.
Sergeant, D M, S Andrews, A Farquhar (2005), 'EVIE Project: User Requirements Analysis Report.'

http://www.leeds.ac.uk/evie/workpackages/wp2/evieWP2_UserRequirementsAnalysis_v1_0.pdf.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003), 'On the science of education design studies', *Educational Researcher,* 32 (1), 25-28.

Shaw, Wendy (2001), 'The use of the Internet by academics in the discipline of English literature: a quantitative and qualitative approach', *Information Research,* 6 (2). http://informationr.net/ir/6-2/ws8.html.

Short, Harold and Marilyn Deegan (2004), 'Expert Seminar: What do Humanities Researchers need from ICT?,' Arts and Humanities Research Council ICT Strategy Review. http://www.ahrcict.rdg.ac.uk/activities/expert_seminars/. http://www.ahrcict.rdg.ac.uk/activities/expert_seminars/humanities_needs.pdf.

Silverman, D. (2004), *Doing Qualitative Research: A Practical Handbook*. London: Sage.

Smith, N., Ferguson, N. and Schmoller, S. (2004), 'Personalisation in Presentation Service.' www.therightplace.net/jp/.

Spärck-Jones, K. (2005), 'E-resources for Research in the Humanities and Social Sciences - A British Academy Policy Review.' http://www.britac.ac.uk/reports/eresources/index.html. http://www.britac.ac.uk/reports/eresources/report/eresources-pdf.pdf.

Sparks, S. (2005), 'JISC Disciplinary Differences Report.' http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc.

Stephen, T. and Harrison, T. (2002), 'Building Systems Responsive to Intellectual Tradition and Scholarly Culture ', *The Journal of Electronic Publishing* 1. http://www.press.umich.edu/jep/08-01/stephen.html.

Stewart, D.W., and Shamdasani, P.N. (1990), *Focus Groups: Theory and Practice*. Newbury Park, CA: Sage.

Stone, S. (ed.), (1980), 'Humanities Information Research: Proceedings of a Seminar: Sheffield 1980' (Centre for Research on User Studies, 4; Sheffield: British Library Board).

Talja, S. (2002), 'Information sharing in academic communities: Types and levels of collaboration in information seeking and use', *New Review of Information Behaviour Research,* 3, 143-59. http://www.uta.fi/~lisaka/. http://www.uta.fi/~lisaka/Taljaisic2002_konv.pdf.

Unsworth, J. (2003), 'The Humanist: "Dances with Wolves" or "Bowls Alone"?' paper given at Scholarly Tribes and Tribulations: How Tradition and Technology Are Driving Disciplinary Change, October 17. http://www.arl.org/scomm/disciplines/Unsworth.doc http://www.arl.org/scomm/disciplines_program.html.

University of Bristol Information Services (2004), 'Portals and Portal Frameworks.' www.bris.ac.uk/is/projects/portal/portalbytes/.

Wagner, E. and Newell, S. (2004), "Best' for Whom?': the tension between 'best practice' ERP packages and diverse epistemic cultures in a university context', *Journal of Strategic Information Systems*, www.elsevier.co/loacate/jcis.

Waloszek, G. (2001), 'Portal Usability – Is There Such A Thing?' *SAP Design Guild,* Edition 3. http://www.sapdesignguild.org/editions/edition3/overview_edition3.asp.

Warwick, C., Terras, M., Huntington, P. and Pappa, N. (2006) 'If you build it will they come? The LAIRAH Study: Quantifying the use of online resources in the Arts and humanities through statistical analysis of user log data'. School of Library, Archive and Information Studies, University College London

Watry, P. and Larson, R. (ed.), (2005), 'Cheshire 3 Framework White Paper: Implementing Support for Digital Repositories in a Data Grid Environment, Local to Global Data Interoperability - Challenges and Technologies, 2005.' http://cheshire.berkeley.edu/Cheshire_Sardinia.pdf#search=%22%20%22cheshire%203%20framework%20white%20paper%22%22.

Watson P. (2003), 'Databases and the grid' In Berman et al 2003a pp 363 – 384.

Weymouth, T., Hardin, J., Golden, G., Severance, C. and Leasia, J. (2003), 'CHEF at the University of Michigan-and Onward', paper given at CIC Workshop.

http://www.cic.uiuc.edu/groups/CIOTechForumPlanningCommittee/archive/ConferencePresentati
on/TechForum2003/Topics/CollaborationTools_Weymouth.pdf#search=%22chef%20and%20the
%20university%20of%20michigan%20and%20onward%22.

White, C. (2003), 'Is the Portal Dead?' *DM Review*.
http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=6959.

Wilson, A. (ed.), (2004), '2003 OCLC Environmental Scan: Pattern Recognition' Dublin, Ohio: Online
Computer Library Center.
http://www.oclc.org/membership/escan/downloads/research.pdf.

Wilson, A. (1987), 'Libraries In Support of Scholarly Communications in the Humanities', Occasional
Publication of the University of London Library Resources Coordinating Committee, 7.

Working Group on Research Infrastructure in the Humanities and Social Sciences (2004), 'Blueprint for the
European Resource Observatory for the Humanities (EROHS).'
http://www.erch.info/Default.aspx?ID=3.

Wouters, P. (2004), 'The Virtual Knowledge Studio for the Humanities and Social Sciences @ the Royal
Netherlands Academy of Arts and Sciences',
http://www.virtualknowledgestudio.nl/en/
http://www.niwi.knaw.nl/nl/initiatieven/digitalisering/vks/C%3A%5CDocuments+and+Settings%
5CElly%5CMy+Documents%5CThe+Virtual+Knowledge+Studio+for+the+Humanities+
and+Social+Sciences.pdf.

Wright, C. (2006), 'AHRC ICT Strategy Project: Peer Review and Evaluation of Digital Resources for the
Arts and Humanities.'

Wynne, M. (2002), 'Linguistics and the Arts and Humanities Data Service.'
http://www.llas.ac.uk/resources/conferenceitem.aspx?resourceid=1314.

--- (2002), 'Setting the Agenda: Languages, Linguistics and Area Studies in Higher Education: 24-26 June
2002, Manchester Conference Centre, UMIST.'
http://www.ilas.ac.uk.

--- (2004), 'Evaluation in the Arts and Humanities Data Service', Journal of Information and Knowledge
Management Systems, 4.
http://eprints.ouls.ox.ac.uk/archive/00000826/.
http://eprints.ouls.ox.ac.uk/archive/00000826/01/eprint.pdf.

Young, R. (2004), 'Findings of the Altis Survey 2003-2004.'
http://www.portal.ac.uk/spp/documents/testing/phase2/altis/AltisSurvey20032004.doc

Zaphiris, P. (2004), 'Usability Studies -JISC Services and Information Environment.'
http://www.jisc.ac.uk/uploaded_documents/JISC-Usability-Studies-Final.doc.

A1: The RePAH Team

A2: Arts and Humanities Research Communities

A3: The Development of Portal Provision in the Arts and Humanities, 1996-2006

A4: WP1: The Online Questionnaire and its Results

A5: WP2: Web-Log Analysis Report

A6: WP3: First Focus Groups

A7: WP4: Analysis of the Delphi Analysis

A8: WP5: The Managed Research Environment Demonstrator

A9: WP6: Phase II User Trials of the Managed Research Environment Demonstrator

A10: WP7: A Review of Intute Functionality

A11: RePAH Project Activities and Outputs

Appendix A1 The RePAH Team

Knowledge Media Design
De Montfort University
Portland 2.3a
The Gateway
Leicester LE1 9BH

Professor Stephen Brown	0116-257-7173	sbrown@dmu.ac.uk
Mr. Robert Ross	0116-207-8161	rross@dmu.ac.uk
Mr. David Gerrard	0116-250-6382	dgerrard@dmu.ac.uk

The Humanities Research Institute
University of Sheffield
34, Gell St
Sheffield
South Yorkshire S10 2TN

Professor Mark Greengrass	0114-222-2574	m.greengrass@sheffield.ac.uk
Mr. Jared Bryson	0114-222-9896	j.bryson@sheffield.ac.uk

Appendix A2 The Arts and Humanities Research Community

Prepared by Mark Greengrass

2.1 The Arts and Humanities Community by Discipline

AHRC PANELS	RAE 2001 UOA	RDN PORTAL/Intute – categories as indicated in the top-level Intute-Humanities resource list	AHDS PROVISION [partial provision indicated in square brackets]
Panel 1: Classics, Ancient History and Archaeology	57 [Classics, Ancient History, Byzantine and Modern Greek Studies] 58 [Archaeology]	Historical and Philosophical Studies [Archaeology | History | History and Philosophy of Science Philosophy | Religion and Theology]	[AHDS-Archaeology]
Panel 2: Visual Arts and Media: practice, history and theory	33 [Built Environment] 60 [History of Art, Architecture and Design] 64 [Art and Design] 65 [Communication, Cultural and Media Studies]	Arts and Creative Industries [Architecture | Communications, Media and Culture | Design Fashion and Beauty | Music and the Performing Arts | Visual Arts]	[AHDS Visual Arts]
Panel 3: English Language and Literature	45 [American Studies] 50 [English Language and Literature]	Literature, Linguistics, Classics [Classics | Comparative Literature | English Studies | Linguistics]	AHDS Literature, Language and Linguistics
Panel 4: Medieval and Modern History	59 [History] 45 [American Studies]	Historical and Philosophical Studies [Archaeology | History | History and Philosophy of Science Philosophy | Religion and Theology]	AHDS History
Panel 5: Modern Languages and Linguistics	46 [Middle Eastern and African Studies]	European Languages, Literature, Historical and Cultural Studies [Celtic | French | German | Hispanic | Italian |	AHDS Literature, Language and Linguistics

	47 [Asian Studies] 48 [European Studies] 49 [Celtic Studies] 51 [French] 52 [German, Dutch and Scandinavian Languages] 53 [Italian] 54 [Russian, Slavonic and East European Languages] 55 [Iberian and Latin American Languages] 67 [Linguistics]	Portuguese | Russian Scandinavian | Slavonic and East European | Other European Modern Languages (General)] Other Languages, Literature, Historical and Cultural Studies [African | American | Australasian | Chinese | Japanese Latin American | Middle Eastern | South Asian | Other Asian]	
Panel 6: Librarianship, Information and Museum Studies	61 [Library and Information Management]	General Arts and Humanities [Cross-disciplinary (Arts) | Humanities (General) Humanities Computing | Manuscript Studies Museum/Library/Archive]	
Panel 7: Music and Performing Arts	66 [Drama, Dance and Performing Arts] 67 [Music]	Arts and Creative Industries [Architecture | Communications, Media and Culture | Design Fashion and Beauty | Music and the Performing Arts | Visual Arts]	AHDS Performing Arts
Panel 8: Philosophy, Law and Religious Studies	62 [Philosophy] 36 [Law] 63 [Theology, Divinity and Religious Studies]	Historical and Philosophical Studies [Archaeology | History | History and Philosophy of Science Philosophy | Religion and Theology]	

Figure 1

2.2 Arts and Humanities Researcher Populations Based on 2001 RAE Reports

	AHRC Subject Panel 2001 RAE Subjects Below	Totals No. Individuals Submitting to RAE		Institutional Size By Staff Numbers/ Total Institutions With Departments		
1	**Classics Ancient History and Archaeology** 1 Classics, ancient history, Byzantine and Modern Greek Studies—346.6 2 Archaeology—482.5	829.1	1	50+ 30-50 10-30 -10	1 1 15 9	26
			2	50+ 30-50 10-30 -10	1 2 18 5	26
2	**Visual Arts and Media: practice, history and theory** 1 Built Environment—600.5 2 History of Art, Architecture & Design—346.5 3 Art & Design—1669.5 4 Communication, Cultural & Media Studies—358.6	2975.1	1	50+ 30-50 10-30 -10	1 1 29 6	37
			2	50+ 30-50 10-30 -10	-- -- 14 25	39
			3	50+ 30-50 10-30 -10	5 12 35 19	71
			4	50+ 30-50 10-30 -10	-- -- 15 23	38
3	English Language and Literature	1519.6		50+ 30-50 10-30 -10	2 7 50 30	89
4	Medieval and Modern History 1 History—1077.9 2 American Studies—113.5	1833.4	1	50+ 30-50 10-30 -10	2 10 49 32	93

				2	50+	--	13
					30-50	--	
					10-30	4	
					-10	9	
5	Modern Languages and Linguistics 1 Middle Eastern & African Studies—128.9	2158.4		1	50+	--	11
					30-50	--	
	2 Asian Studies—129.5				10-30	5	
					-10	6	
	3 European Studies—558.7			2	50+	--	13
					30-50	1	
	4 Celtic Studies—92.1				10-30	3	
					-10	8	
	5 French—446.1			3	50+	1	38
					30-50	1	
	6 German, Dutch & Scandinavian				10-30	19	
	Languages—255				-10	17	
	7 Italian—103.5			4	50+	--	15
					30-50	--	
	8 Russian, Slavonic, and East European				10-30	2	
	Languages—77.3				-10	13	
	9 Iberian and Latin American			5	50+	--	43
	Languages—157				30-50	1	
					10-30	20	
	10 Linguistics—210.3				-10	22	
				6	50+	--	40
					30-50	--	
					10-30	6	
					-10	34	
				7	50+	--	19
					30-50	--	
					10-30	1	
					-10	18	
				8	50+	--	17
					30-50	--	
					10-30	--	
					-10	17	
				9	50+	--	31
					30-50	--	
					10-30	5	
					-10	26	
				10	50+	--	24

				30-50	--	
				10-30	10	
				-10	14	

6	Librarianship, Information and Museum Studies
	    1 Library and Information
	    Management—302.1
		302.1	1	50+
				30-50
				10-30
				-10
					--
					--
					13
					10
						23

			2	50+ 30-50 10-30 -10	--	

7	Music and Performing Arts
	    1 Drama, Dance and Performing
	    Arts—395.9
	
	    2 Music—486.3
		882.2	1	50+
				30-50
				10-30
				-10
					--
					1
					13
					26
						40

			2	50+ 30-50 10-30 -10	-- -- 18 41	59

8	Philosophy, Law and Religious Studies
	    1 Philosophy—460
	
	    2 Law—1352.9
	
	    3 Theology, Divinity and Religious
	    Studies—439
		2251.9	1	50+
				30-50
				10-30
				-10
					--
					1
					22
					21
						44

			2	50+ 30-50 10-30 -10	3 15 28 14	60
			3	50+ 30-50 10-30 -10	-- -- 14 29	43
	Arts and Humanities TOTALS	12751.8				

Figure 2

2.3 Unit of Assessment Size in RAE2001

UoA Size in RAE2001	Numbers of Units Recorded
<10	474
10-30	430
30-50	53
>50	16

Figure 3 [Source: HERO]

2.4 HE Higher-Degree Qualifications obtained in the UK in the Arts and Humanities [2004-5]

Subject Area	Total Higher Degrees	Doctorates
Architecture, Building and planning	**2910**	**240**
Law	**5785**	**200**
Mass communications and documentation	**3245**	**75**
Languages	**5520**	**895**
Historical and Philosophical Studies	**4740**	**925**
Creative arts and design	**5030**	**275**
TOTALS	**27,230**	**2610**

Figure 4 [Source: HESA]

2.5 Sector Distribution of UK Professional Archaeologists in c.2000

SECTOR	Distribution [%]
Archaeological Contractor	**30.4**
Local Government	**18.1**
Heritage Agencies	**15.5**
University HEI and Research Groups	**14.7**
National Museums	**3.6**
Independent Consultants	**3.5**
Archaeological Societies	**1**
Other central government funded organizations	**0.1**
Other organizations	**13.2**

Figure 5 [Source: Condren, Richards, Robinson and Wise (1999)]

2.6 Taxonomy of Knowledge Structures

Group	Knowledge	Culture
Physical Sciences e.g. Physics	Cumulative, atomistic (crystalline/tree-like); concerned with particulars, qualities, simplification; resulting in discovery/explanation	Competitive, gregarious; politically well-organised; high [publication rate; task oriented
Humanities e.g. history	Reiterative; holistic (organic/river-like); concerned with particulars, qualities, complication; resulting in understanding/interpretation	Individualist, pluralistic; loosely-structured; low publication rate; person-oriented
Applied Sciences, e.g. mechanical engineering	Purposive, pragmatic (know-how via hard knowledge); concerned with mastery of physical environment; resulting in products and technologies	Entrepreneurial; cosmopolitan; dominated by professional values; patents substitutable for publications; role-oriented
Applied Social Sciences, e.g. education	Functional, utilitarian (know-how via soft knowledge); concerned with enhancement of (semi-) professional practice; resulting in protocols and procedures	Outward-looking; uncertain in status; dominated by intellectual fashions; publication rates reduced by consultancies; power-oriented

Figure 6 [Source: Sparks (2005), following Fry (2004)]

2.7 Relationship between degree of 'mutual dependence', 'task uncertainty', and the production and use of digital resources – Three Exemplary Fields

Exemplary Field:	- High energy Physics	Corpus-based Linguistics	Social/cultural Geography
Dominant Culture	High degree of mutual dependence, with low degrees of task uncertainty	Moderate degree of mutual dependence with moderate degree of task uncertainty	Low degree of mutual dependence with high degree of task uncertainty
Differential role of Informal and Formal Communication	Speedy establishment of knowledge claims via informal communication; system of conference papers and pre-prints; publication mainly serves citation criteria	Need to communicate a high concentration of technical information supported in conference proceedings, reports and manuals	Formal communication system unimportant because of lower levels of interpersonal recognition (e.g. low people to problem ratio) and need to justify goals, approaches and techniques in literature; informal communication system determined by individual groups and specific social networks
Role of ICT in Communication Systems	Tightly coordinated system for the informal dissemination of research results via integrated digital networks; production of centralised field-based digital resources	Quest for the development of coordinated systems for the informal dissemination of research results hindered by local ICT infrastructures; decentralised locally-produced field-based digital resources	Non-production of field-based digital resources; reliance on commercially-produced generalist digital resources@ ICT infrastructure determined at the level of the employing institution rather than the field or discipline

Figure 7
[Source: Fry (2004)]

2.8 Information-Seeking Behaviour of Arts and Humanities Scholars

1. 'Essential Resources'
[Q: What is the single most essential resource you use, the one that you would be lost without?]

[results in % within each group]	Medical and biological sciences	Physical sciences and engineering	Social Sciences	Languages and Area Studies	Arts and Humanities
Pre-Prints		5.8	1.4		1
Post-prints		6.3	9		3.9
Journal Articles	90.7	71.6	69.3	28	27.2
Conference proceedings		5.8	5		
Books	.6	1.4	9.2	50	35.9
Datasets	4.3	3.4	7.8	2	2.9
Technical reports		1			
Government or NGO reports	1.2		2.3		
Legal Sources			.5		
Other Textual Sources			3.7	10	14.6
Non-Textual Sources	.6		.5	2	8.7
Other	2.5	4.8	4.1	8	4.9
TOTAL	100	100	100	100	100

2. Search-tools and Sources of reference
[Q: 'What search tool or reference source is most essential to you, the one you would be lost without?']

[results in % within each group]	Medical and biological sciences	Physical Sciences and engineering	Social Sciences	Languages and Area Studies	Arts and Humanities
Other	13	5.7	6.7	8	3.9
Subject-specific abstracts and indexes	18.5	20.6	22.4	6	13.6
Subject-specific online gateways	22.8	3.3	6.7	2	2.9
General Bibliographic resources	9.9	11.5	15.2	46	29.1
Citation Databases	21	21.5	9.9	4	3.9
Search Engines	14.8	36.4	35.9	24	36.9
Works of Reference		1	3.1	10	9.7
TOTAL	100	100	100	100	100

1. Informal Resources

[% of respondents in each group]	Asking a Colleague	Emailing a colleague or peer	Reading email newsletters	Posting an enquiry to an email list	Reading blogs
Medical and biological sciences	80.2	87	17.9	11.7	4.3
Physical sciences and engineering	81.9	81.9	21.9	12.4	4.3
Social Sciences	76	78.2	35.6	15.1	7.1
Languages and area studies	74	80	16	12	2
Arts and Humanities	76.7	76.9	31.1	21.4	6.8

4. Problems in Accessing Research Resources

[% of respondents in each group	Medical and biological sciences	Physical sciences	Social Sciences	Languages and Area Studies	Arts and Humanities
Library does not take the journals I need	89.4	82	81.9	79.2	69.1
Library does not buy the books I need	18.8	31.5	38.1	62.5	61.8
Library does not subscribe to the databases I need	22.4	36	34.3	16.7	32.7
I cannot get access to the conference proceedings I need	18.8	44.9	25.7	12.5	23.6
Key information is proprietary	10.6	12.4	17.1	8.3	12.7
I need to travel to access resources and funding isn't available	9.4	14.6	24.8	58.3	58.2

Figure 8 [Source: Sparks (2005) – based on 750 completed questionnaires from individuals divided into subject groups by 2001RAE UoA]

2.9 'Road-Map' of Arts and Humanities Research Activities
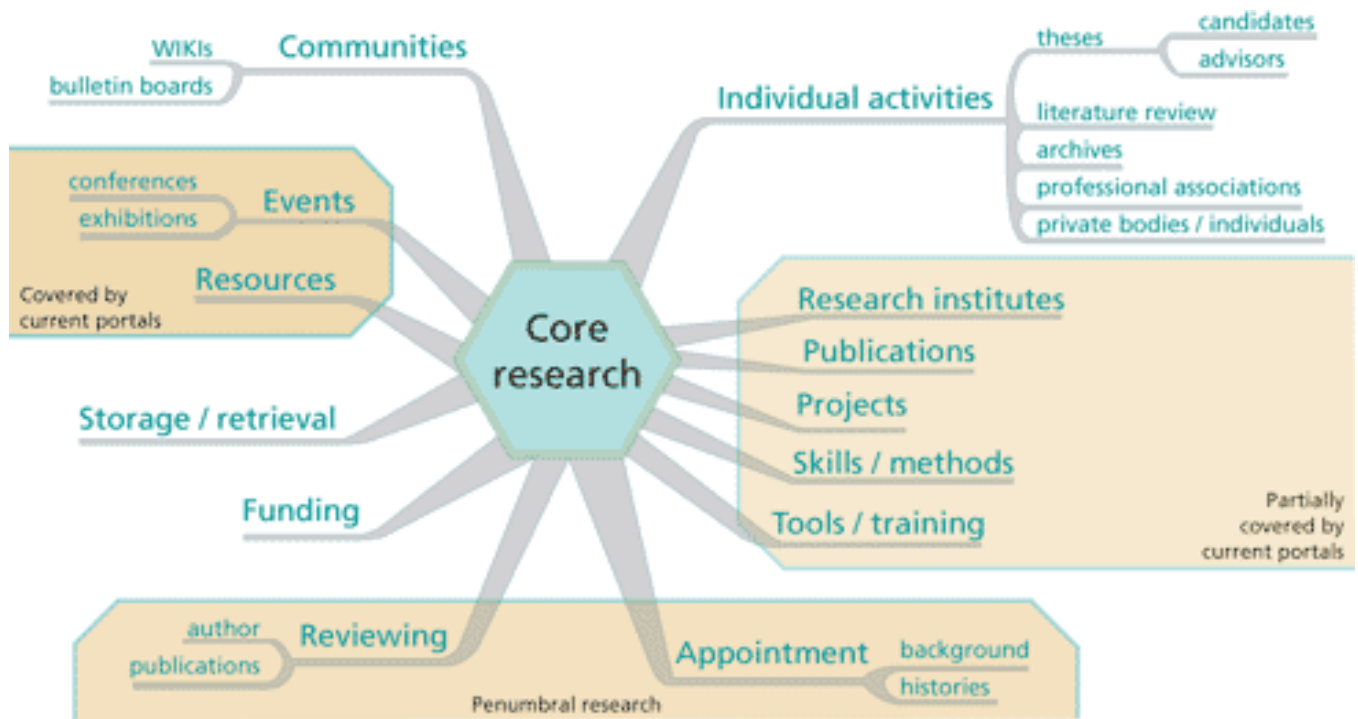
Figure 9

Appendix A3 The Development of Portal Provision in the Arts and Humanities, 1996-2006

Prepared by Mark Greengrass

3.1 The Concept of e-infrastructure

The infrastructure of academic scholarship has developed over centuries. For the Arts and Humanities, that means its institutional fabric – libraries, archives, museums, research centres, etc. It also means the tools of scholarship – bibliographies, searching aids, concordances and editions, journals and academic presses – that make information accessible. The equivalent infrastructure for academic scholarship is also needed for electronic media. It is often referred to as 'e-infrastructure'(UK) or 'cyberinfrastructure' (US), meaning (as the American 'Atkins Report' defined it) the 'middleware' that links base technologies with specific software programmes, services, instruments, data, etc in a now widely-understood framework:

- o a **baseware** 'layer of **base technologies**…the integrated electro-optical components of computation storage, ad communication'
- o a **middleware** 'layer of **enabling hardware**, algorithms, software, communications, institutions, and personnel' that lie between'
- o a **topware** layer of 'software programs, services, instruments, data, information, knowledge, and social practices applicable to specific projects, disciplines, and communities of practice.'

The UK has been at the forefront over the past decade in developing its middleware e-infrastructure for the Arts and Humanities. This has been under the auspices of the Resource Discovery Networks and the Arts and Humanities Data Archive. The latter was singled out in the July 2006 US Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences as an international exemplar of e-infrastructure in this area ['Our Cultural Commonwealth,' 2006]. This Appendix presents the evolution of those bodies, as they relate to the Arts and Humanities, over the past decade.

The evolution of e-infrastructure in the Arts and Humanities reflects the exponential growth of the WWW from 1995 onwards. The initial middleware concept of the 'portal' rapidly gathered pace in the late 1990s, reaching a climax in around 2000, coinciding with the dot-com boom (c1997-2000). From 2000 onwards, technological developments refined the portal concept, offering more complex and interactive portal frameworks. Since that date, the alternatives for harvesting, managing, accessing and publishing information within organisations have also developed rapidly, leading to the growth of institutional portals, sometimes referred to as 'special interest', 'vertical' or 'niche' portals. So, too, has the sophistication of the interface with the WWW. Commercial search tools (Google: Yahoo: About: Go.com: Lycos, etc) developed portal services, sometimes referred to as 'horizontal' or 'mega-portals', beyond their traditional search tools in their competition to be a 'starting-point' of choice, aggregating information in order to keep people at their site and draw repeat visitors [Lamb, 2004]. Libraries and academic institutions have been relatively quick to see the advantages of using a single digital interface for a variety of administrative and teaching functions.

3.2 The Evolution of RDN Subject Portals

The RDN Subject Portals began life as subject **'gateways'**, a term that gained currency in the UK through the Electronic Libraries Programme, funded by the JISC following a call in August 1994 [eLib, 2004]. The underlying concept emerged in response to the challenge, as it then appeared, of 'resource discovery' in the rapidly growing Internet environment. These initially took shape from 1994 onwards in the context of the Access to Networked Resources [ANR] component of the eLib Programme. Following consultation and a bidding process, a number of subject 'gateways' were established or funded, based on recommendations of the Access to Networked Information Resources [INIR, 1993], commissioned in 1993. One prototype subject 'gateway' was already in the process of development since the ESRC had funded a project in the summer of 1992 to assist UK social scientists in the use of networked information. SOSIG, as it became known, went live in the summer of 1994 with a descriptive environment for about 300 Internet resources. The subject 'gateways', funded by the JISC in 1994 and operational from 1995/96, were:

    SOSIG [Social Science Information Gateway]
    EEVL [Edinburgh Engineering Virtual Library]
    OMNI [Organised Access to Medical Networked Information]
    HISTORY [a gateway for History]
    ADAM [Art, Design, Architecture and Media Information gateway]
    BIZ/ED [A gateway for business studies, economics, accounting, leisure, sport & recreation and travel & tourism]

The ANIR report accurately reflects the dominant priorities of the period:

- o the need to create 'access' and 'discovery' services.
- o the belief in centralised provision of such services as an emerging 'academic infrastructure' within a relatively coherent UK higher educational framework.
- o an awareness that subject 'gateways' were dependent on the development of broader technical standards and protocols in an area where there was considerable fluidity and unpredictability.
- o an aspiration to influence the evolution of technical standards through creating centres of activity with sufficient critical mass to establish a consensus.

It should be noted that Humbul was in existence as early as 1986 based at the University of Leicester's Office for Humanities Communication, and operated as a bulletin board on the JANET network for computing in the humanities, including people, events and publications [McCarty, 1989; Fraser, 2006].

3.2.1   In retrospect, it is difficult fully to recapture the discussions and environment which led to the formation of the 'subject gateways'. The web was not yet an overwhelmingly predominant network environment in 1994; and the network itself did not have the range and pervasiveness that it would soon develop. Gopher (1991) permitted the construction of user-orientated and browse-able services and Mosaic (1993) provided a browser. But no single access protocol allowed users to reach all resources of interest and, for a time, there were different, albeit often interconnected, resource spaces

in existence (Gopher: HTTP: WAIS: ftp, etc). Subject 'gateways' provided a 'resource discovery' service that was badly needed in certain academic domains for three perceived reasons:

i) *subject taxonomies and ontologies*. It was recognised at an early stage that it was not practicable with the search environments then available to browse through highly populated resources organised in a flat and undifferentiated way. Structures began to be introduced to divide up the resource domains by access method, geographical area and by subject delineations. Some attempts were made to adapt library classification systems to facilitate browsing by subject.

ii) *more elaborate searching mechanisms*. These were developed in order to complement browsing. They provided increasingly important navigational tools for large resource spaces. But the reliability of the search tools was dependant on the terse, non-descriptive texts from which the indexes to the materials were created. This prompted the investigation of enriched resource descriptions, delivered in database-driven services or as resource guides. The literature from this period includes schema for resource descriptions and templates (e.g. Internet Anonymous ftp Archive [IAFA] templates) that offered the potential for a service which contained full enough descriptions to allow the user to assess a resource without having to retrieve or connect to it initially, but not so full or complex as to require a lavish outlay of cost and very specialised staff to create. These IAFA templates were widely adopted by the eLib gateways, where they were used in association with the WHOIS++ protocol in ROADS servers. In the late 1990s there was an aspiration that the practical experience of the subject 'gateways' from 1995 onwards could influence the development of this format. By contrast, the literature contained little by way of realisation of the development of other metadata formats (Dublin Core being the most powerful candidate to support resource descriptors) which might be automatically harvested by the subject 'gateways'.

iii) *quality controlled environments*. The literature of the period from 1995 to 2000 reflected the awareness that, in contrast to the print environment, the network environment had no established quality control mechanisms such as pre-publication peer-review, the recognised 'brand' of a well-known publishing house and its series, and post-publication peer review processes. Engagement with the subject communities through workshops tended to suggest that there was potential value in a moderated collection of resources, managed to ensure a level of quality and collected to ensure a level of relevance. How those levels of quality and relevance were to be assessed, however, remained unclear.

3.2.2   By 1999, three issues were of growing concern in the evolution of the subject gateways'.

i) *Market penetration*. The first published user-evaluation study appeared, based on a small sample of academic users in two universities [Mackie and Burton, 1999]. It concluded that the gateways were positively welcomed by some members of the academic community, but that the majority of academics in the relevant subject communities were totally unaware of them.

ii) *Growth.* There was an awareness of the impact of the overall rapid increase in UK internet bandwidth (up to 2.5Gps by 2001) and also the increasing number of subscribing HEI and FE institutions to the JISC (up from 151 in 1991 to 1,000 in c.2001). So although there was a concern about the long-term sustainability of funding subject 'gateways' to levels that would guarantee their effectiveness in a teaching and research context, this was overlaid by the technical possibilities that seemed to be emerging and the likely benefits of scale from further investment in resource discovery.

iii) *More powerful ways of linking distributed electronic resources.* The emergence of common metadata standards and structured protocols offered the technical prospect of assembling and linking resources in a way that was concurrently being implemented at the RDN in hybrid libraries. This was seen as a way of counteracting the emerging problems of separate cataloguing formats within the existing subject 'gateways'.

3.2.3   In 1999, the JISC took the strategically important decision to establish the DNER [Distributed National Electronic Resource]. One of the earliest initiatives of the DNER was to found the Resource Discovery Network [RDN]. A contract to run the RDN was awarded to King's College, London, with UKOLN at the University of Bath as a partner with responsibility for technical interoperability as between the various 'gateway' providers]. The DNER programme began with the notion of moving from a 'gateway' to a 'hub'. These hubs were established around broad faculty-wide subject divisions and embraced the pre-existing subject 'gateways'. Their initial focus was to provide a 'secure and convenient access to a range of information services and resources' through a 'web-based front-door' [Pinfield and Dempsey, 2001; JISC 2002a; JISC, 2003b]. The following 'hubs' were created or emerged in 1999-2000, each established in leading institutions in order to create a more sustainable structure:

EMC [engineering, maths, computing, embracing EEVL] – Heriot-Watt University

BIOME [health, life, and biomedical sciences, embracing OMNI] –University of Nottingham

SOSIG [social sciences, business, law] – Bristol University

Humbul [begins hosting the RDN site for the Humanities in August 1999] – University of Oxford

PSIGATE [newly created for the library and information sciences] – University of Manchester

A Maths Portal for the mathematical sciences, based at the University of Birmingham, was separately funded. In addition, a consultancy was initiated to advise about provision of a 'hub' for the Creative Arts and Industries. The RDNC Consultancy Report was one of the few exercises in this period to include an evaluation of potential user needs.

3.2.4   The RDN/JISC decision reflected new strategic thinking. Broader subject domains were chosen to facilitate partnership, sustainability and preserve existing investments. The 'hubs' were expected to take the initiative in establishing domain-specific services. The model was designed to provide alternative possibilities for developing a critical mass of resource descriptions across a broader range of subject

areas. At the same time, the 'hubs' were asked to provide additional functionality to access distributed network resources. A dominant aspiration was the creation of a more highly interconnected information and learning environment to support UK learning, leaching and research. A workshop held early in the life of the RDN in 1999 to discuss issues of business planning. It concluded that there were various funding patterns across the new 'hubs'. Some of them had commercial partners. Others were part of a wider service, offered by Research Councils. The rest stood alone.

3.2.5   The years from 2000 to 2005 were marked by the **patchy and uneven development.** By 2005 the following features were integral to Humbul's service development:

- o cross-searching from one 'hub' to another
- o user profiles
- o user-authentication to access distributed information
- o brokering services, providing cross-searching of distributed materials consistent access to bibliographic resources (e.g. the integration of serials article locator services)
- o alerting services via email
- o information feeds

There was a programme during this same period known as the Subject Portals Project (SPP). It arose from a one year programme entitled 'Subject Access to the DNER' (SAD) funded by the JISC in November 2000, and concentrating on the technical aspects of developing the RDN 'gateways' into 'portals'. Under SAD I, this followed by a second phase also funded by the JISC. Some work was done on collection development (identifying those collections to which access would be provided by the particular portal), building a Z39.50 cross-search prototype on the SOSIG, EEVL and BIOME 'hubs', some work on user-profiling, and some portal design development. This was followed by a larger, second phase [SAD II], also funded by the JISC, which took place in 2002 and 2003, by which time it had become the 'Subject Portals Project'. A 'Phase II' of the 'Subject Portals Project', funded by the JISC, then began in 2003 and completed its work in February 2005. This was an ambitious agenda, and in retrospect it seems that the complexity of the task was underestimated. By that date, the Subject Portals Project had still failed to live up to its ambitious expectations. One should bear in mind, however, that this was a period when **new technical standards** for portal development were emerging. The Java portlet standard JSR 168, and the Web Services portlet standard WSRP only became defining standards for allowing different portlets and portal frameworks to interoperate in the course of 2004-5. Beta-test sites of particular software developments in various areas were produced on particular subject gateway platforms and open-source code was made available. By that date, the user of Humbul (the more established and advanced of the two Humanities Portals) had access only to the following additional 'portal' services:

- o Provision of RSS news feeds relating to Humbul database content
- o Provision of third-party RSS feeds (but only jobs.ac.uk provided)

- o Email alerting service
- o User profiling and improved saved search functionality
- o Web-based Directory of relevant e-Journals

The following features were activated only within the Subject Portals Project environment and not made available more generally in Humbul:

- o Cross-searching of remote arts and humanities databases
- o Integration of ATHENS single sign-on system to enable access and searching of remote databases

The following features were tested within Humbul, but not activated or not taken forward:

- o Harvesting and indexing of third-party OAI metadata
- o Provision of Z39.50 service (provided for a time, but then taken out of operation)
- o Provision of public OAI metadata repository (provided for a time, but taken out of service)
- o Provision of an Events database (proposed for inclusion but incomplete development)
- o Harvesting and indexing of online peer-reviewed ejournals (proposed but not implemented)

From February 2005 onwards, development work concentrated on the provision of the core subject-wide 'information discovery portal', now launched as Intute. The Intute portal is mainly designed as an integrated 'portal' across the whole disciplinary spectrum [http://www.Intute.ac.uk/about.html]. Arts and Humanities are branded as one of four main subject areas, with pre-existing subject domain categories retained in the migration. Although it was not part of our remit, we have included some investigations of the Intute-Arts and Humanities as it impacts on the recommendations in this report as an additional work-package within our research programme [**A8].**

At the launch of Intute in July 2006, Humbul and Artifact databases contained around 18,000 publicly available records. Approximately 11,680 of these were from the former Humbul database, where a further 3,685 records were in a mixture of draft, suspended or queued records being processed.

3.3 User-Requirements Analysis and RDN Arts and Humanities Developments

The RDN Arts and Humanities e-infrastructure provision during the period from 1996-2006 was **more influenced by a hard-science model of information needs** and driven by **technical developments** as much as by **perceived discipline requirements**. We have located four surveys of information needs undertaken in connection with the RDN programme:

> 1) A 1996 preliminary survey of user information needs and search needs undertaken by Alison Ferry to inform the design of ADAM gateway in art, design, architecture and media studies [Ferry, 1996]. It was based on 723 completed responses to a distributed questionnaire.

2)   A preliminary user-survey undertaken for SAD-1 [Guy, 2003]. User scenarios were developed 'in-house' by SOSIG, BIOME and Humbul to help to scope the requirements for the planned alerting and user-profiling services.

3)   A series of user-consultation exercises were undertaken by each of the 'hubs' [Subject Portals: Phase One Documents, 2003].

4)   A more general survey of portal functionality undertaken by ALTIS, information scientists and specialists at the University of Birmingham and a part of the RDN [Young, 2004].

These surveys are of purely historical interest now, and we have not sought to compare them directly with our own evidence. Those in the period up to 2000 tended to be more orientated towards librarians and other information gateway managers and potential middleware providers. Some were more explicitly targeted towards teaching needs. Most of the user-requirements analysis was limited to testing 'functionality' and 'usability' of a particular feature that had already been envisaged or developed. So, the SAD-1 developed 'user scenarios' for SOSIG, BIOME and Humbul. Only two focus groups were held, both within the engineering domain and EEVL.   These assisted in the development of user-requirement specifications for the SPP events and aggregated news services. Usability testing was conducted initially on internal subject portal staff. A small sample of user were invited to undertake nine specified tasks to familiarize themselves with the workings of the particular 'hub' portal, and then asked the following questions:

- o   What do you like about the portal?
- o   What don't you like?
- o   What should work better?
- o   Would you use the portal for your own research?
- o   Would you use it in preference to a general search engine such as Google?

The Humbul user-consultation involved a small focus-group (3 undergraduates; 3 postgraduates; three library staff; and 1 lecturer) whose discipline backgrounds were not specified [Subject Portals: Phase One Documents, Humbul User Testing Report, 2003]. The exercise concentrated on an evaluation of 'hub' functionality. In the answers to the five more general questions, the following user-evaluation issues were raised:

- o   The interdisciplinary component of Humbul was appreciated. Cross-searching 'could be very popular' but 'whether I would ever use that I do not know'.
- o   Screen layouts and search facilities were variously interpreted. Some thought that they were 'cluttered' and 'not intuitive enough', 'confusing' with 'too many options and technical language'. Others appreciated an 'excellent research facility', but one that required familiarization by users to be 'decoded'
- o   The more practical elements of Humbul's delivery received the warmest praise. These included 'jobs.ac.uk', the 'search landscape tool' and the 'storage system'. The elements most criticised were the taxonomies of the resource descriptions ('resources need to be described in a better way when listed'), some vagaries of behaviour in the search engine, some distaste for the side newsbar, and some difficulties with the save and print functions

- o The utility of the site for research purposes was not universally appreciated. Some thought they would 'definitely' or 'probably' use it. They appreciated the 'tailored' humanities approach that it afforded. Others thought that it might only be of use for 'general research' and that it was not ideal for 'more specific research'

Humbul had more functionality than several respondents expected. Several participants had no prior knowledge of what a 'portal' might achieve. There was only a small range of usability that was identified specifically as not currently being provided (a 'way of narrowing searches'; 'forums; help; friendly introduction'; 'picture-only search facility'). It was compared unfavourably to JSTOR, LION and the then new Web of Knowledge. In comparison with GOOGLE, opinions were more divided. Humbul was 'clearly much better for academic purposes', 'more complex', 'better organised' and 'far better in terms of relevance'; but GOOGLE was 'simpler'.

How, if at all, these evaluations fed through to modifications in the design, presentation and functionality of the RDN portals is unclear.

For the purposes of e-infrastructure development, the more recent and general survey of portal functionality undertaken by ALTIS, information scientists and specialists at the University of Birmingham is of some relevance [Young, 2004]. The survey covered all the disciplines of the RDN, and was conducted from 1 December 2003 to 4 January 2004 via its web site. It attracted 243 respondents. Each of the following eight questions were scored a value from 1 (low) to 5 (high), depending on how the user felt about the statement:

- o I mainly use Google search the web.
- o It would be good to see a variety of news sources in one location.
- o I would like to see a conference and events listing.
- o Email alerts of new resources would be something I would like to see.
- o I would use a service that searches multiple databases from one location.
- o Personalisation of a website is something I would use if it had the right options.
- o A full text journal search is something I would use.
- o I like web site interaction and enjoy being involved (e.g. forums, surveys)

The most striking conclusions were the ubiquity of Google as a web-search tool, and the more mixed responses to email alerts, news feeds and conference and events listings. Although these were generally viewed positively, there was an understandable hesitancy about being inundated with material not directly relevant to one's interests.

3.4 The Evolution of the AHDS

3.4.1   The Arts and Humanities Data Service is not strictly comparable to the RDN information gateways. It is a service-provider, established to 'collect, preserve and promote' electronic materials resulting from research and teaching in the arts and humanities. Its mission statement is to serve the arts and humanities education community by:

- o Preserving arts and humanities digital resources created by Higher Education
- o **Providing rich, deep, access to the intellectual content of arts and humanities digital resources created by and for Higher Education**
- o Supplying advice and guidance in the creation of digital resources to quality standards that ensure their suitability for informed use in research and research-led teaching, and their long-term viability [AHDS Strategic Plan 2002-5]

But resource discovery was regarded as an essential complement to its collections development from its inception. In this report, we shall be concentrating on that element of its activities, whilst being aware that it is an infrastructural service with a major role in other areas.

The AHDS was established in 1996 as a result of three specialist consultation exercises. The first, prepared by the British Library, concentrated on the expanding horizons for the application of information technology to humanities scholarship [Information Technology, 1993]. The second examined the conservation, curation and resource discovery issues from a library perspective [Report, Funding Councils' Libraries Review Group, 1993]. The third, commissioned by the Information services sub-committee of the JISC, furnished a prospectus, institutional framework and outline methodology, drawing on the model of the Social Science Data Archive, which had been formed four years previously [Burnard and Short, 1994]. From its inception, it was conceived as a distributed service, made up of five service providers (two of which were in existence prior to 1996), held together by an Executive, based at King's College, London [Greenstein and Trant, 1996]. Initially, each separate provider held its resources independently of the other. But, shortly after its inception, the AHDS began pioneering the possibility of exploiting resource discovery metadata, using the Dublin Core as an interchange format and Z39.50 as a network application protocol standard [Miller and Greenstein, 1997] By 1998, the AHDS 'gateway' provided its collections catalogues as a virtual uniform catalogue. At the same time, this catalogue permitted users to register with the AHDS, to acquire access to its holdings, to save queries between sessions, and to access a list of AHDS resources suited to their own resource discovery requirements. In its advanced search form, it was also possible to search for other (i.e. non-AHDS) online information resources in any query [Greenstein, 1998]. Because of the wide variety in its holdings and the different disciplines it serves, there was no attempt to implement any controlled vocabularies in the resource descriptors. So, e.g. Anglo-American Cataloguing Rules are appropriate to, and adopted by, AHDS Literature, Language and Linguistics, whereas the Art and Architecture Thesaurus provides the controlled vocabularies in use by AHDS Visual Arts. Further development work therefore took place on a common metadata framework, based around the RSLP (Research Support Libraries Programme) Collection Development Schema [RSLP, 2006]. This was then mapped onto the five existing collection-level metadata schemas in order to permit more detailed search options [Anderson, 2004]. The new cross-search catalogue was launched in October 2003. In 2004, there were changes in nomenclature that reflected the greater coherence of the service and the growing role for the Executive of the service. Otherwise, the basic structure of the AHDS has remained stable until the present [Dunning, 2004]:

- o   Archaeology Data Service – now AHDS Archaeology [York]
- o   History Data Service – now AHDS History [Essex]
- o   Oxford Text Archive – now AHDS Literature, Language and Linguistics [Oxford]
- o   Performing Arts Data Service – now AHDS Performing Arts [Glasgow]
- o   Visual Arts Data Service – now AHDS Visual Arts [Farnham]

Its role as a curator of electronically-created materials was substantially enhanced by the decision of the Arts and Humanities Council in 1999 to require funded projects which produced electronic content to deposit it with the relevant AHDS service.

3.4.2 The pattern of collection growth within these service-providers, as recorded in the AHDS Annual Reports, supplemented by individual service-provider Annual Reports (where available), reflects fundamental particularities in the way in which the disciplines they serve have responded to the application of information science:

AHDS Archaeology began life in October 1996. Its activity reflected, from an early stage, archaeologists' extensive reliance upon computer techniques. Archaeology Data Service Annual Reports have been analysed from 1996-7 through to 2004-5 [Archaeology Data Service Annual Reports]. They present a detailed picture of a well-organised service that has developed a good understanding of its client communities' needs. It has a large (over 50) Advisory Committee and, from its first year, organised expert workshops and regular liaison meetings with its practitioner-base. From early on, it also cultivated collaboration with the numerous local, regional and national agencies that develop and maintain the UK's archaeological record. This is reflected in the 139 collections currently available for search in ArchSEARCH. They include (to highlight, by way of example, some of the major distributed national collections for which it serves as an important resource discovery gateway for its community) the Defence of Britain Archive (databases from field and documentary work carried out between April 1995 and December 2001), the CBA reports (a complete series of Council for British Archaeology Research Reports), its links to the English Heritage National Inventory (NMR) and Index to Microfilmed Archaeological Archives, and the Society of Antiquaries Library Catalogue. It has significant relationships with the Natural Environment Research Council (NERC), archiving some of the digital data produced in that field. It also has relationships with English Heritage through RECAP (Rescue of Completed Archaeological Reports) [Anderson, 2004, p.3], and with developer-funded archaeological projects, such as the Channel Tunnel Rail Link [CTRL]. The AHDS is responsible for 234 archaeology-related collections funded by the AHRB/C and the British Academy. It is also responsible for 150+ collections funded by other public and commercial funding bodies.

AHDS History was founded in January 1993 as a specialist unit within the United Kingdom Data Archive [UKDA] at the University of Essex. Its resource discovery function has, from its inception, been subsumed (at least to some extent), within this very significant gateway to major government datasets of economic and social statistic surveys (including the census), and an even wider range of international economic and social

statistic datasets, generated by world bodies such as the OECD, IMF, IEA and World Bank. We have examined its Annual Reports from 1999-2000 to 2004-2005. They provide an impression of an institution that is offering a wide range of services for data creators, depositors, researchers, teachers and the wider community. It has traditionally relied on working relationships with professional bodies (the *Association for History and Computing UK*; the *Social Science History Association*, etc) to keep in touch with its client communities, along with a small Advisory Board, attendance at conferences, and a small range of expert workshops. Its substantially-used resources include longitudinal studies (e.g. the National Child Development studies from c.1960s onwards), and a substantial range of qualitative datasets, mainly from the late nineteenth and twentieth centuries (e.g. 'Family, Life and Work Experience, 1873-1973'), census data and the 'Historic Parishes of England and Wales'. The Great Britain Historical Database brings together a very considerable range of census and other data from the later nineteenth century onwards. At the same time, it hosts a more disparate, but substantial, range of pre-contemporary datasets and deposited material. The number of datasets consulted has significantly risen in recent years – from 163 in 2003-4 to 254 in 2004-5, or a third of its collection by title [AHDS Annual Report, 2004-5, p. 11]. It now has an aggregated collection of 627 'studies' (the UKDA equivalent of collections). There are some legal issues regarding the organizations identified in the licence form that dictate where the collections can be hosted. For this reason, the physical hosting of its collections is divided between the AHDS Executive in London and the UKDA.

AHDS Literature, Language and Linguistics grew out of the Oxford Text Archive, established as part of the Oxford University Computing Service 30 years ago. It does not appear to have developed a strategy of relating to its client communities. Traditionally, it archived electronic texts of interest not just to literary textual scholars, but to those working in linguistics, law, history and theology. It thus accumulated materials in any literary genre, period or language and, in the past, been a supplier of large-scale digital libraries, electronic text archives and commercial data providers, of which (to some degree) it was a pioneer. in the period from 1976 to 1996, it collected 2081 collections which are currently stored by the OTA but not currently available for download. The licenses for these collections was signed with the University of Oxford and the AHDS is therefore unable to take responsibility for them. One consequence of its longer paternity is that only recently have the collections ingested there since the inception of the AHDS begun to be transferred to the AHDS shared repository, a process that had not yet been completed by the summer of 2005 [AHDS Annual Report, 2004-5, pp. 12-3]. It has been faced with different methodologies and varying standards for defining and creating text corpora [Wynne, 2002]. Its most requested resource is still, apparently, the Toronto Dictionary of Old English, originally deposited in 1985, a reflection of the rapidly increasing significance accorded to computer-applications in linguistics. In the period from 1996 to 2006, it ingested 433 collections.

AHDS Performing Arts focuses on collecting digital resources across the broad field of the performing arts – music, film, arts, theatre, broadcast arts, and dance. It is currently hosted by the Humanities Advanced Technologies and Information Institute (University of Glasgow). We have only located one published Annual Report for this service (2002-

3). That confirms our impression of a service that has had difficulty establishing itself, defining its mission and relating to its client community. There is no mention of any Advisory Group and no apparent strategy of being able to take into account user needs. This is particularly significant since its base-community is broad, and the disciplines within it relatively 'immature' in academic terms, especially in respect of the creation and scholarly use of digital materials. In addition, this is an area where the relevant applications are technically more sophisticated and make more demands upon arts research practitioners. There are substantial copyright restrictions, partly reflecting the finance and culture of the performing arts, to materials in this area [Anderson, 2004, p. 3]. In addition, the creators of resources in the performing arts have often invested heavily in the created of a 'value-added front-end' to their resource that cannot easily be transferred to the AHDS. Music; and Film, Television and Radio Studies are areas in which significant resources were made available first, followed by Theatre and Dance. The online distributed database to collections of music materials in the UK (CECILIA) is an example of techniques developed in other AHDS service providers being successfully cross-fertilised to the arts area. In 2004-5, four new collections were accessioned, and a further three converted for delivery. But in 2005-6, a further 12 were foreshadowed [Anderson, 2005, p. 14]. It is now responsible for a total of 32 collections with a further 4 in various stages of processing.

AHDS Visual Arts was launched in March 1997 and is now based at the University College for the Creative Arts (Farnham Campus). It serves an area in the arts where there are more digital collections than for the performing arts. Many of them arise, however, from the galleries, museums and heritage sectors. So, like Archaeology, this service provider has had to develop sophisticated collaborative relationships in the development of its searchable collections. The National Fine Art Digital Collection (www.fineart.ac.uk) is one example – a prototype searchable catalogue dataset of 11 fine art collections, curated by UK HEI and consulted by over 160,000 visitors in 2004-5 [AHDS Annual Report, 2004-5, p. 15]. We have examined its Annual Reports from 1997-8 through to 2004-5 [AHDS Visual Arts, Annual Reports]. Perhaps because the Visual Arts DS initially had a consortium structure (made up of four constituent organisations) the impression is of a service that had placed a particular emphasis on relating to its user-base from its inception. It has a large Advisory Group, a tradition of regular workshops, training and small-group functions in different HEI throughout the UK. The range of its collections and links has grown significantly. Seventy-nine new collections were ingested in 2004-5, with 49 of them being made available from the website in 2004-5 [AHDS Annual Report, 2004-5, p. 15]. By the summer of 2005, its image catalogue contained over 50,000 records, and they are of increasing significance for research practitioners in the humanities as well as arts. It now has a total of 105 collections, of which 76 are image collections, 16 are learning and teaching collections, and 11 'other resource' collections.

3.4.3 The growth of the AHDS-curated holdings over the period from 2001/2 to 2004/5 reflects the **differential patterns of development** of the branches of the AHDS, and therefore the way in which their user communities relate to them. The sharp rise in acquisitions in 2004-5 reflects, in part, the impact of the first tranches of resource

enhancement and research grant projects coming to fruition. But it also is the result of the growing maturity of the links between the AHDS and other local and national bodies, and HEI. These are important elements in the user-evaluation environment for the AHDS's resource discovery role:

Figure 1

Total number of new acquisitions

2001-02	**63**
2002-03	**83**
2003-04	**98**
2004-05	**166**

Figure 2
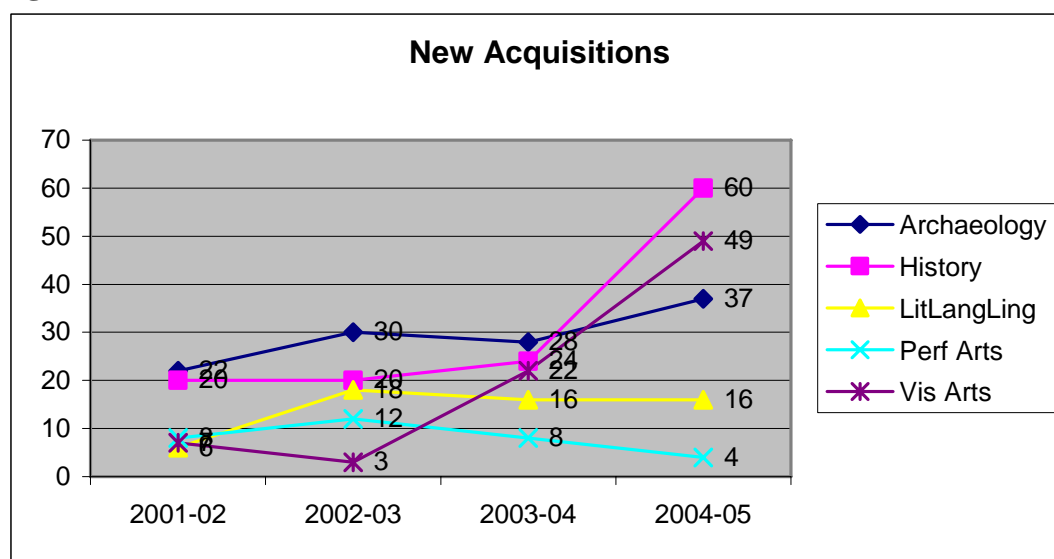
Source: AHDS Annual Report, 2004-5, p. 18

The AHDS is now responsible for a total of 1,225 collections.

3.4.4 User-Needs Evaluation in the context of the AHDS. We have located only a small number of user-requirement studies in relation to the AHDS over the past decade. We are aware of others being mentioned, but we have no documentation for them [e.g. Greenstein, 1998]. They are:
- o A user-needs survey was conducted by the Visual Arts DS in December 19997-February 1998, based on a paper and online questionnaire, to which it had 107 responses [Grout and Rymer, 1998]
- o A user-needs survey conducted by the Archaeology Data Service in 1999 on behalf of the Digital Data in Archaeology Survey of User Needs Project Consortium

Although the Director of the AHDS acknowledged as early as 1998 that 'how users actually exploit the Gateway, particularly in relation to their use of underlying Service Provider catalogues, will provide useful feedback for the system's further development',

such feedback mechanisms do not seem to have been systematically put in place [Greenstein, 1998, p. 11].

3.5 Other Relevant User-Needs Requirements Analyses

The scope of our study has not permitted us to review the evidence of user-needs studies across the board, not even the **proliferation of 'portals'** that has occurred over this period. One of the dominant trends of the period from 1996-2006 has been the proliferation of websites attached to learned societies and specialist institutes of learning which proclaim themselves as 'portals'. Most of these are, at best, 'thin' portals, offering manually-created pages of news, information and links linked to a local search engine. They typically do not harvest information electronically, or enable the user to do so. They provide no additional services to the user beyond those available from a good internet search-engine. At the same time, there has been the parallel and widespread development of institutional HEI 'portals', often serving as tools for managing the complex protocols for accessing different levels of intranet information as well as a gateway to other externally purchased information providers and gateways. We have done our best to gain a general appreciation of these trends, taking particular note of the report on E-resources for research in the humanities and social sciences prepared for the British Academy in 2005 by Karen Spärck-Jones [Spärck-Jones, 2005]. For more general institutional portal developments (often known in commercial organizations as 'enterprise portals'), we have relied on the Nielson-Norman Group Report of 2005 [Goodwin, Schwartz and Nielson, 2005]. This establishes 'best portal-development practices' on the basis of commercial experience, emphasizing the importance of a portal to provide 'usable information', and therefore regularly matched against 'the needs of users' [p. 15].

3.5.1 A Model in User-Requirements Evaluation

We signal, however, one particular disciplinary area in the Arts and Humanities that provides a model for taking user-needs into account in developing its portal services. HEIRPORT, the Historical Environment Information Resources Portal is the creation of HEIRNET, the Historic Environment Information Resources Network (HEIRNET), and it provides the major portal provision for archaeologists [http://www.britarch.ac.uk/HEIRNET]. HEIRNET is a consortium composed of various public bodies (AHDS Archaeology; the Council for British Archaeology, the Royal Commission on Ancient and Historic Monuments in Scotland, etc) and it has been funded at various stages by the British Library, the JISC, the E-Science Programme, Re-Source, and the National Electronic Library for Health. In 1998, in collaboration with English Heritage and the Royal Commission on the Historic Monuments of England, they commissioned a user-needs analysis for electronic information gateway provision in the sector, which was undertaken in the spring and summer of 1998. It was based on 3,000 questionnaires, mailed to archaeologists and followed up by a smaller number of structured interviews conducted in July 1998. Its focus was on the creation, archiving, use and re-use of digital data in archaeology [Condron, Richards, Robinson and Wise, 1999]. It was a broad-ranging, strategic review, and undoubtedly had a significant impact

in developing service provision in that area. HEIRNET subsequently undertook a further user-evaluation survey in 2002, commissioned from the Cultural Heritage Consortium [Heirnet, 2002]. This was reinforced by a subsequent project which investigated the user-profiles of all the major historic environment information systems over a one-month period in Autumn 2004 and a major User Survey, commissioned by the British Council of Archaeology in 2005 [Brewer and Kilbride, 2005]. Taken together, these surveys have enabled the archaeology community to define its needs, and to see them met, in a way that is unmatched in the rest of the Arts and Humanities sector. HEIRPORT now constitutes the most richly populated (in terms of resources accessible through it) and one of the most elaborate (in terms of attached services) portals in UK Arts and Humanities. It is a model for other disciplines in the Arts and Humanities to follow.

3.6 RDN and AHDS Financial Support

This report is NOT an evaluation of the service currently provided by the RDN and AHDS. The funding of these services is not part of our remit. Their resources have, however, influenced the kinds and levels of information resource discovery that they offer. We therefore provide the following information as part of the background to the user-needs evaluation that we are conducting.   It has been provided by the services, does not address issues of institutional input and overhead, and should be regarded as providing, at best, 'indicative funding levels':

Figure 3

AHDS Funding (1995-2006)

Academic Year	KCL	JISC	AHRC	TOTAL
1995-6	75,408	500,000		575,408
1996-7	50,000	500,000		550,000
1997-8	60,000	325,000		385,000
1998-9	50,000	200,000		250,000
1999-2000	55,000	499,944	261,383	816,327
2000-2001		615,886	298,000	913,886
2001-2		645,000	305,000	950,000
2002-3		547,213	547,213	1,094,426
2003-4		507,638	507,638	1,015,275
2004-5		523,206	523,206	1,046,411
2005-6		534,528	534,528	1,069,056

Humbul Funding

From 2002/3 onwards, Humbul received £128-135,000 per annum from the JISC, with an additional £16,000 in the academic year 2005-6 to fund requirements gathering work to enable Intute-Arts and Humanities to be better adapted to support the research and teaching community. In addition, the service received a further £50,000 per annum from the AHRC. The recurrent funding level for this service has therefore been more or less frozen at 2002/3 levels.

Artifact Funding

From 2002-03, the first year of operation of the Artifact service, it received core funding from the JISC of between £116,339-£152,355 with the breakdown as follows:

- o 2002-03 £147,500
- o 2003-04 £152,355
- o 2004-05 £143,441
- o 2005-06 £116,339
- o 2006-07 £116,339

3.7 RDN and AHDS: Complementary Resource Discovery Agents

Both the RDN and AHDS have developed a resource discovery component to their mission over the past decade. That complementarity is not currently mirrored in their service delivery.

Both the RDN and AHDS have been in place for a decade, but they have not developed best practice **strategies for being in contact with their user communities**.

The evidence from their own user analyses is that their **services are not as well-known or understood as they ought to be.**

Recent developments, in particular the launch of Intute, indicate that there is an awareness of the emerging importance of what one might more properly call a 'managed research environment' in which the twin issues of **access** and **interoperability** can be fully addressed.

Appendix A4 Work-Package 1: The Online Questionnaire Report

WP1 prepared by Robb Ross

A4.1 Introduction

This report investigates the data collected from the online RePAH questionnaire. The findings are based on the completed questionnaires at 30 April 2006: n=149.

A4.2 Demographics

Overall there was a good spread of respondents from the main categories of researcher:
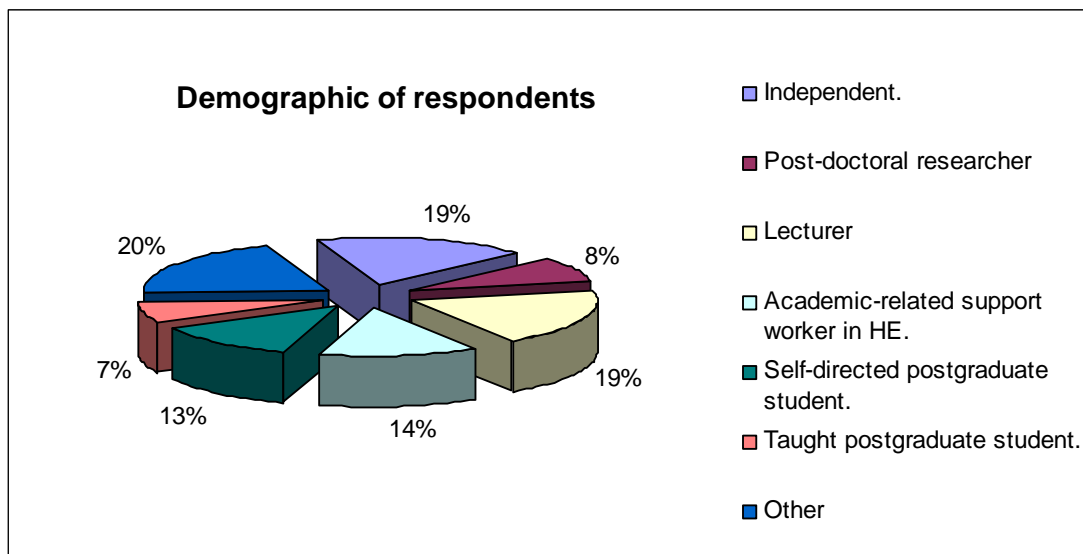
Figure 1: Responses from question 1 "What kind of researcher are you?"

The number of independent researchers did seem excessive but may be a product of the non-UK respondent category (see below). In the 'Other' sector there is a wide variety of respondents; from an emeritus professor to an academic librarian to an IT manager at a research institute.
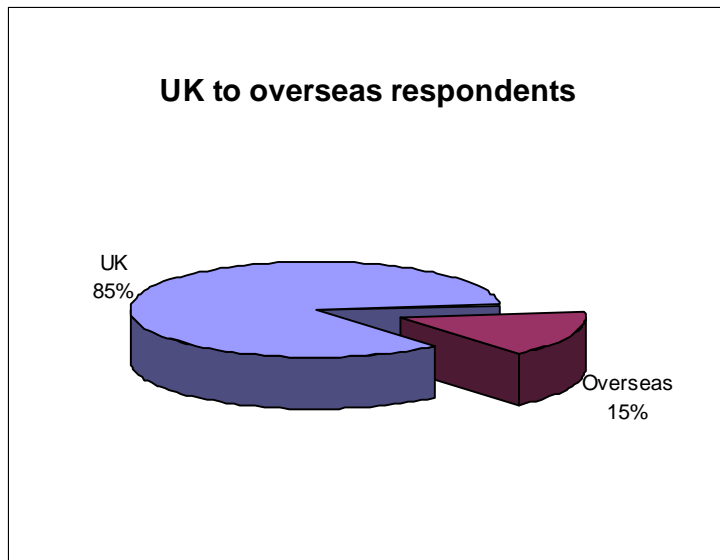
Figure 2: Responses from question 4 "Are you based in the United Kingdom?"

There appears to be a minority number of respondents from outside of the UK. This may be a result of the dissemination exercise including international online newsgroups/electronic newsletters/websites. The largest non-UK respondents were from the United States and Canada, followed by Australia.

The respondents were frequent web users with 89% using the web on a daily basis and 77% have been using the web for 5 years of more:
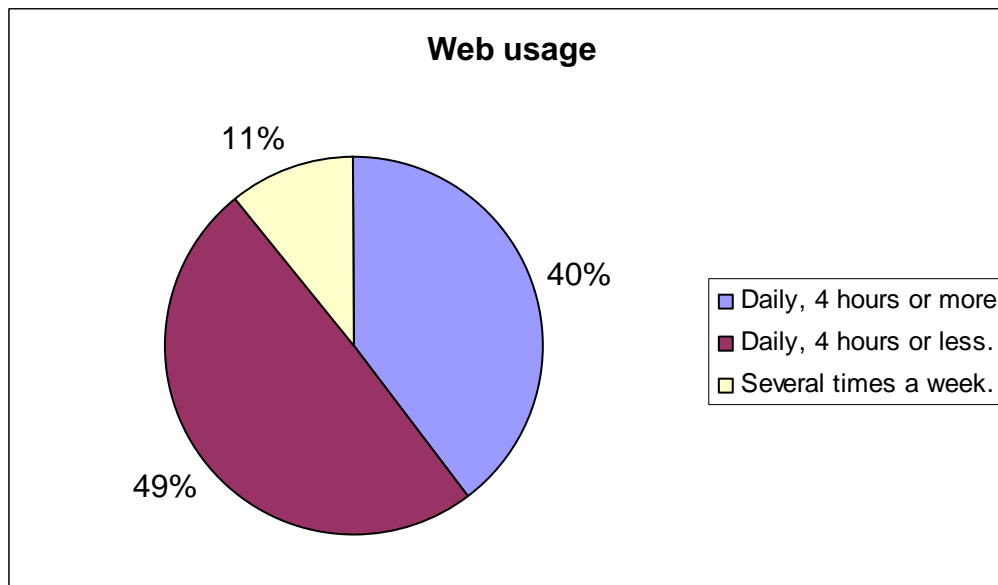
Figure 3: Responses from question 5 "How often do you use the web during the working week?"

Previous research (the Aria project **http://aria.dmu.ac.uk**) identified that many researchers perceive their work to be multi-domain as defined by the AHRC categorisation criteria. In order to capture this more granular information, the design of the questionnaire allowed a choice more than one domain, with the ability to rate more than one as a priority domain.

The resultant research domains to which the respondents said they were affiliated produced some interesting results (see Figure 4). Only 31% of respondents considered themselves to work within a single, significant domain (identified by assignment of the highest rating score), whilst 18% said they had one or more with one individual citing all eight domains. Interestingly 31% of respondents did not consider themselves to have a significant research domain (they did not score any domain with the highest rating). This may be due to respondents rating domains as non-significant due to the researcher's work crossing multiple domains; none of which are considered as primary, although it may also in part be explained by some respondents not being active in research (such as the IT manager).

Further work is required in this area to develop a system whereby researchers can accurately relay how their research fits in to the AHRC domain structure, especially if that research is considered to cut across more than one domain categorisation. However, results from this questionnaire provide adequate information to show, in general terms, the distribution of researchers within each domain.
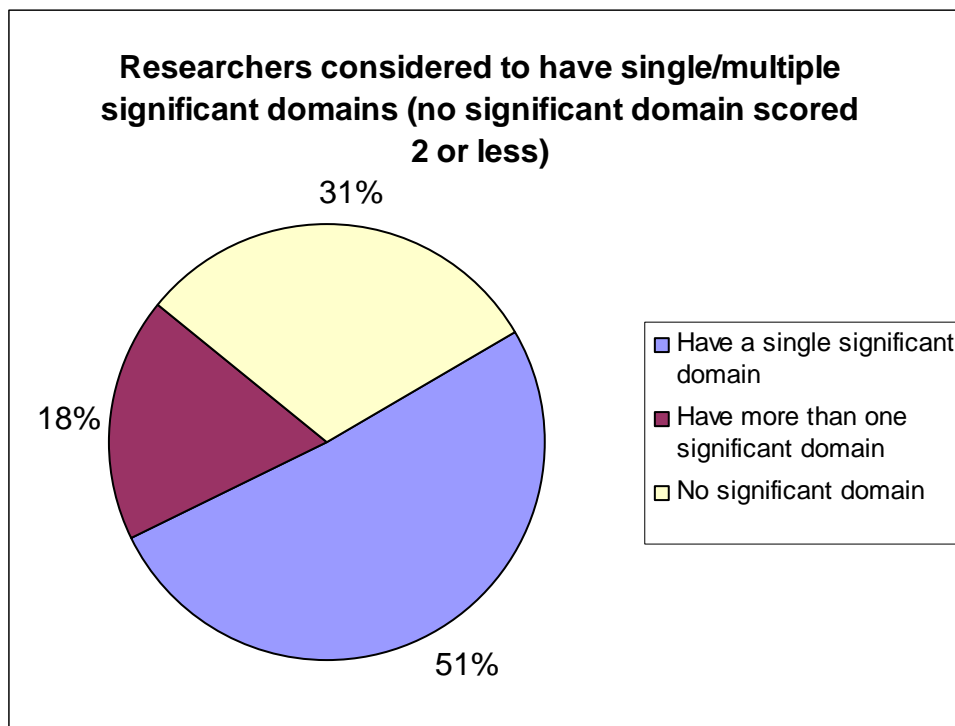
Figure 4: Respondents scoring of AHRC domain significance

Research prioritised by domain

Legend:
- Classics, ancient history and archaeology
- Visual arts and media
- English language and literature
- Medieval and modern history
- Modern languages and linguistics
- Librarianship, information and museum studies
- Music and performing arts
- Philosophy, law and religious studies

x-axis: 1 2 3 4 5 6 7 8

1=High   8 = Low
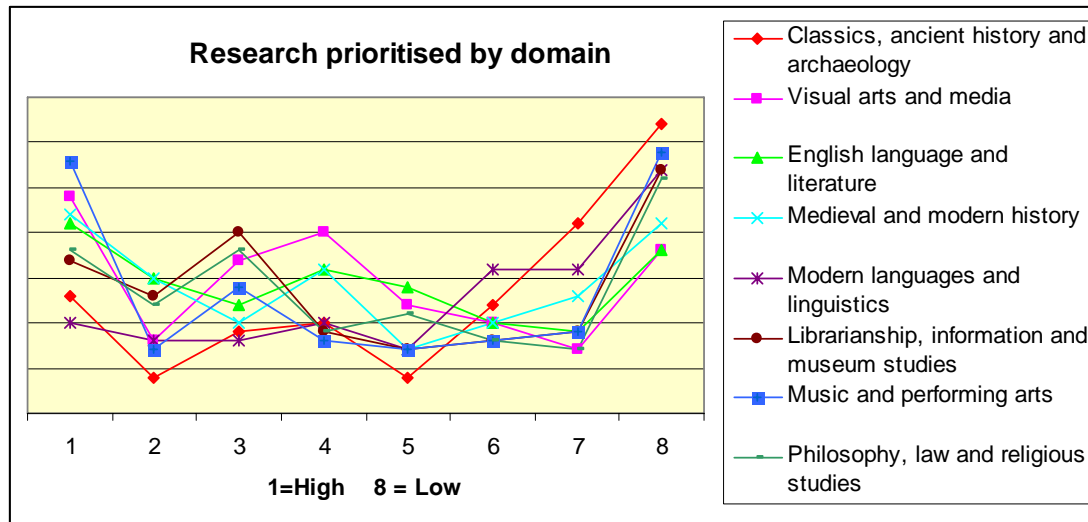
Figure 5: Responses from question 7 "Please tell us what domain the research you undertake falls within."

Whereas it may be difficult for researchers to identify their significant research domain, what can be seen in the results (see Figure 5) is that more people are sure about what domain their research does not cover.

Looking at responses that provided a significant rating for a research domain, the numbers within each are similar, providing comparable responses for each domain.

Respondent domain by significance rating

Legend:
- Classics, ancient history and archaeology
- Visual arts and media
- English language and literature
- Medieval and modern history
- Modern languages and linguistics
- Librarianship, information and museum studies
- Music and performing arts
- Philosophy, law and religious studies
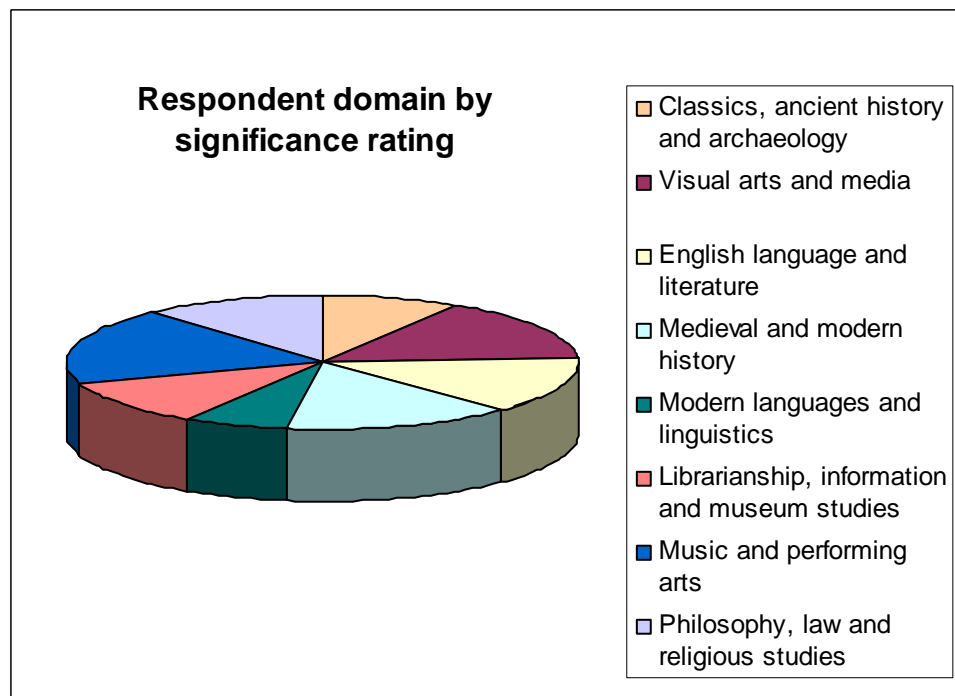
Figure 6: Frequency responses rated by domain significance.

A4.3 Access and use of digital resources

This next section looks at the respondents' access and use of digital resources and how, or if, it has changed their approaches to teaching and research.

The initial question in this section was to ascertain the impact digital resources have had upon current academic working practices.

I could not do my academic work without access to digital resources

61.07%
21.48%
11.41%
4.70%
1.34%
0%

1   2   3   4   5   6

1= Strongly agree   5=Strongly disagree   6=Not applicable
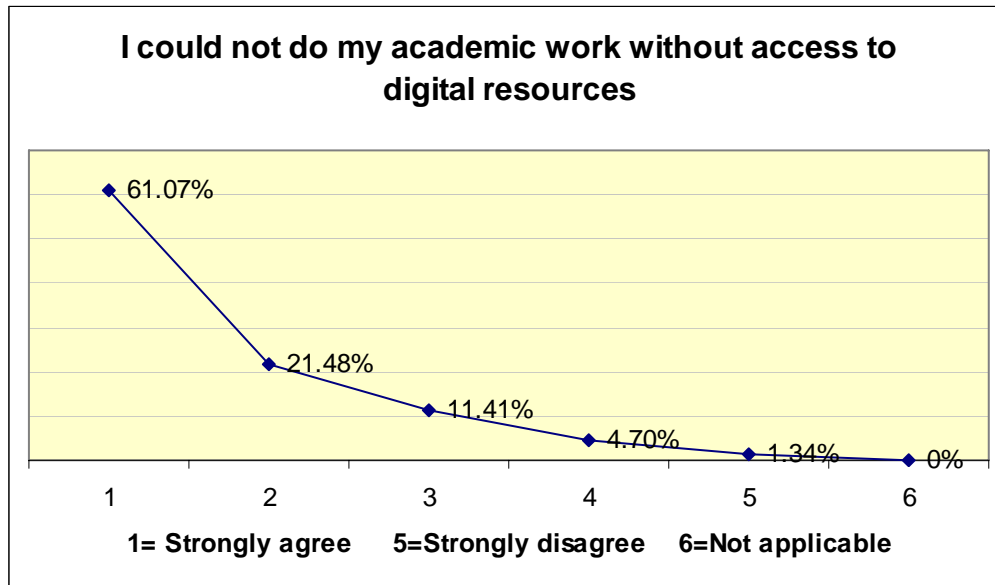
Figure 7: Responses from question 8

The responses show a significant impact upon current academic work and would indicate that digital resources play a crucial role in the ability of researchers to carry out their activities. This is reinforced by the responses to question 10 (see below) which shows that the resources are used extensively and not for a small, but crucial, part of the respondent's work.

I use digital resources extensively in my academic work

- 51.68% (1)
- 29.53% (2)
- 14.77% (3)
- 3.36% (4)
- 0% (5)
- 0.67% (6)

1=Strongly agree   5=Strongly disagree   6=Not applicable
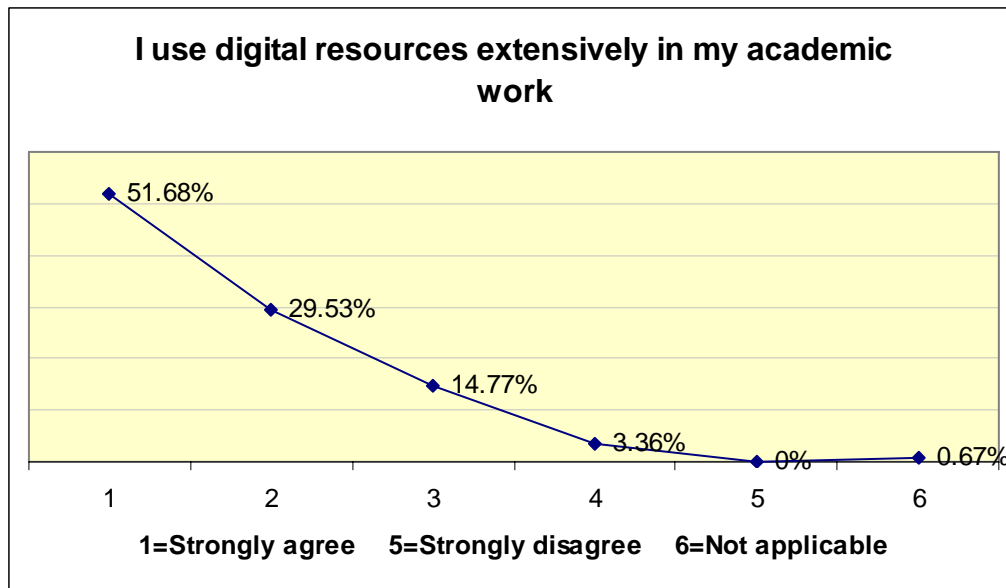
Figure 8: Responses to question 10

Identifying what area of work within which the resources are used shows the resources are used within research activities and not just teaching and that they are of a sufficient quality.

Digital resources are useful for teaching but not for research

- 1.34% (1)
- 4.03% (2)
- 10.07% (3)
- 25.50% (4)
- 48.32% (5)
- 10.74% (6)

1=Strongly agree   5=Strongly disagree   6=Not applicable
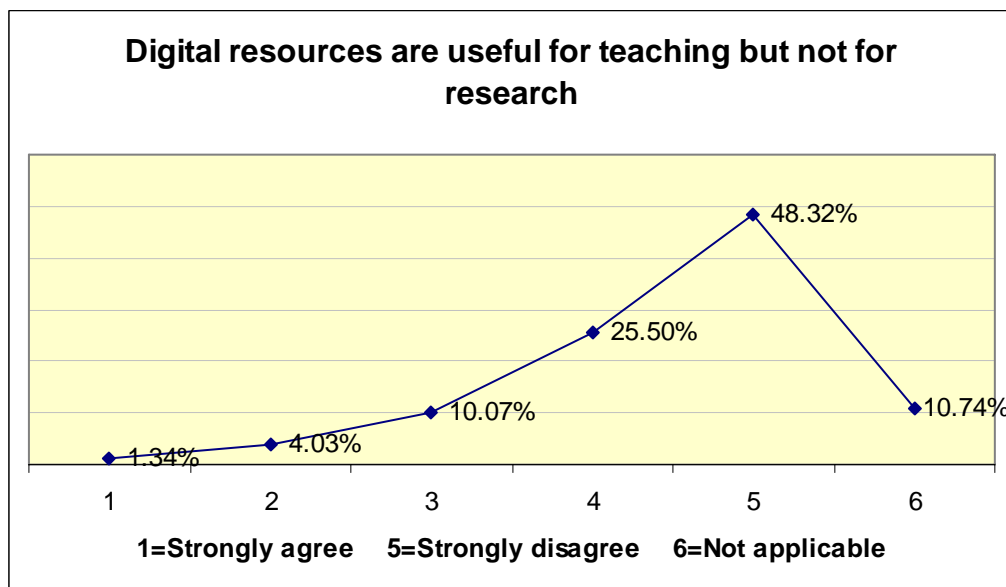
Figure 9: Responses to question 11

The impact that digital resources have made upon working practices is seen as significant by over 40% of research practitioners, with access to scholarly resources via the internet seen by nearly 70% of respondents as one of the major contributors to that work (question 15). Significantly, this type of textual material is not held electronically, but

printed out and read by over 60% of respondents (question 16); although over half of respondents state they prefer digital materials over printed matter (question 18).

Digital resources have changed the way that I do my research

- 40.94%
- 35.57%
- 14.09%
- 4.03%
- 2.01%
- 3.36%

1   2   3   4   5   6

1=Strongly agree   5=Strongly disagree   6=Not applicable
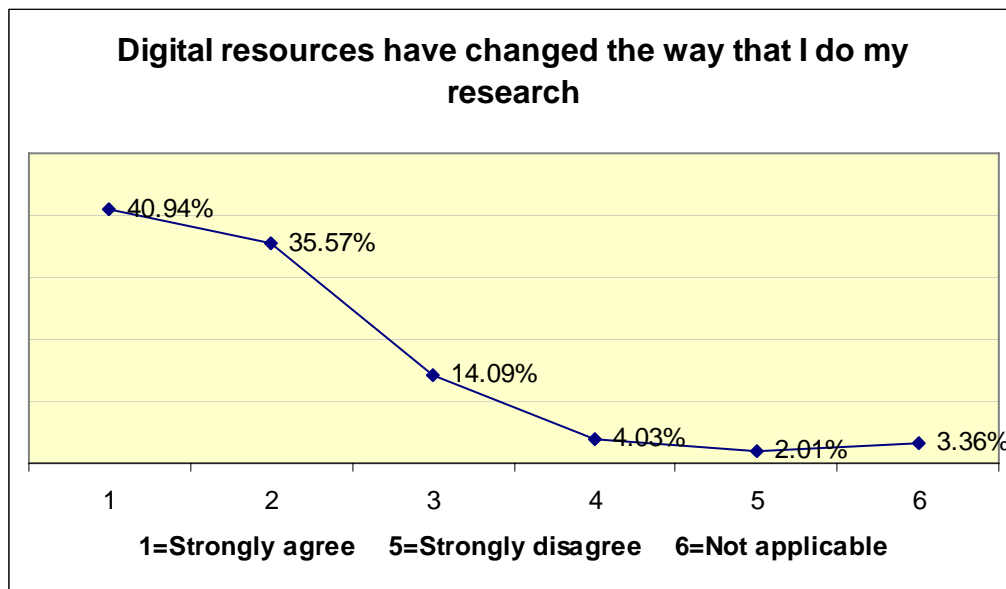
Figure 10: Responses to question 13

Although digital resources have also made an impact upon teaching practices, this has been to a lesser degree or is seen as having no relevance to teaching at all. This is a surprising piece of data considering the increasing use of VLEs, PowerPoint and other digital means of presentation in teaching. Further investigation to identify what users consider digital resources in teaching to be may explain this discrepancy; a misunderstanding as to the meaning of the phrase could have occurred.

Digital resources have changed the way that I teach

- 22.15%
- 23.49%
- 12.08%
- 4.70%
- 1.34%
- 36.24%

1   2   3   4   5   6

1=Strongly agree   5=Strongly disagree   6=Not applicable
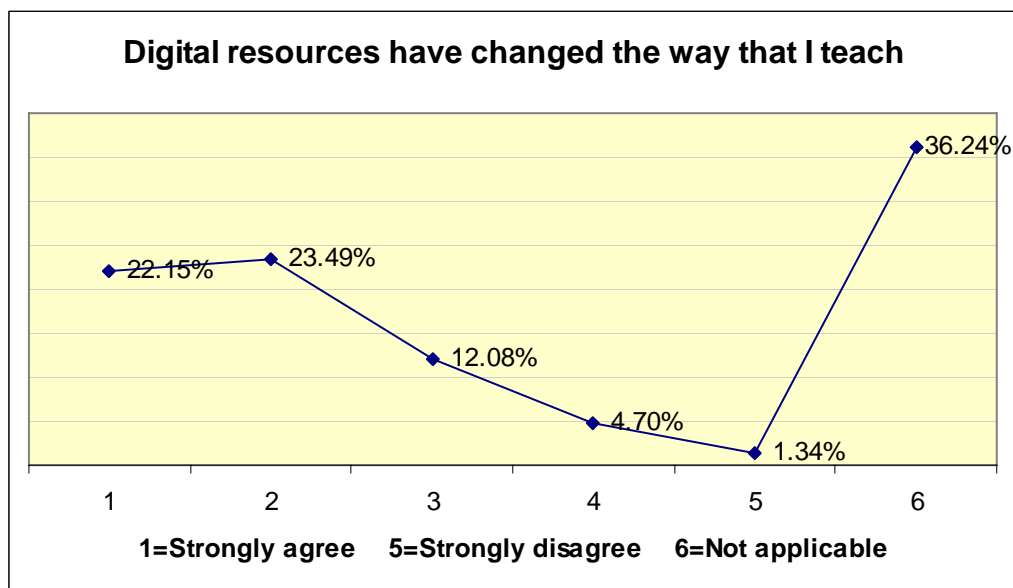
Figure 11: Responses to question 14

When asked about the resources or websites respondents found useful, the most quoted source was University library services such as COPAC or associated access permissions to resources (such as journals) gained from it. The next most quoted resource was Google and its attendant functions such as Google scholar or Google Images with JSTOR and AHDS services the next most quoted. The 'Other' category includes all suggestions that attracted fewer than 6 comments. The use of library services as the main source of information discovery is also supported by a later question (question 41) concerning how the respondents found the questionnaire, with the highest number stating it was found via a library webpage.
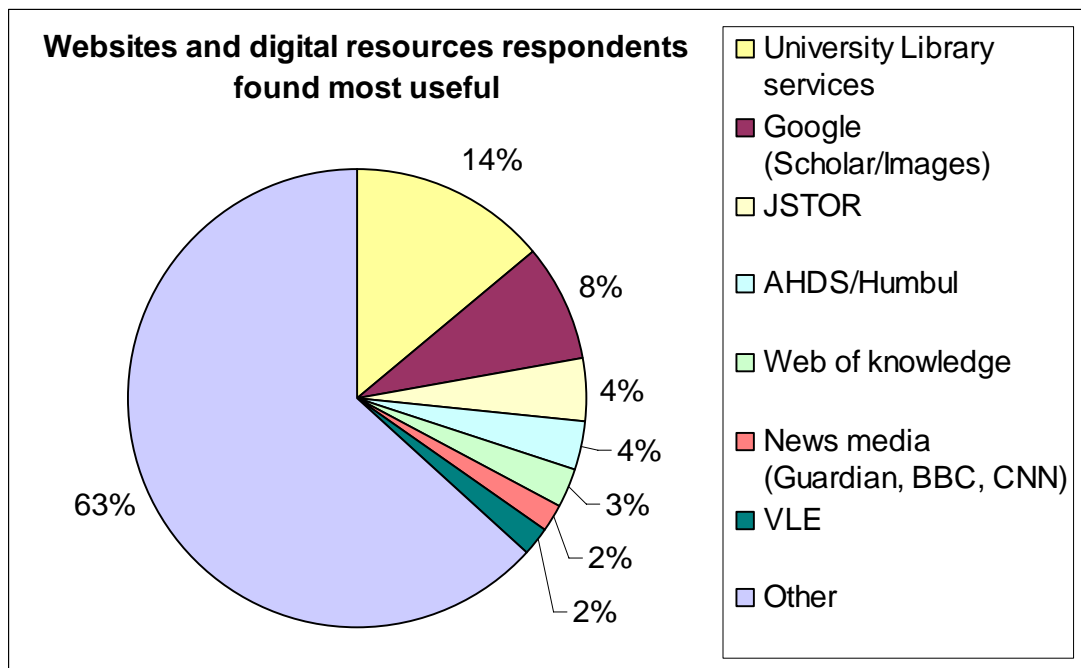
Figure 12: Respondents most quoted digital resources [1]

The free text responses that support this question provided a divergent view of digital resources and access to them. The main point to come out was only allowing three choices was deemed too restrictive, as respondents used far more than this on a daily basis "Three is too few! I have extensive bookmarks and my institution has a portal to a vast range of subscription material". There were also points raised about the access to state-funded electronic resources for those not within the mainstream academic institutions, and the inability for independent researchers to use them "I am extremely distressed at the current trend of privatizing various indexing systems by making them only accessible to institutional subscribers." Finally, there was one comment that was at odds to all others "the internet is very dangerous when using it for research so I believe a book is more resourceful. I only ever use the internet as a last resort for academic work!"

---

[1] A breakdown of the section 'Other' can be found at **A5.7** below.

Breaking the data down into domains, all but classics, ancient history and archaeology and Visual arts and media cite the University Library as their most used digital resource; these two domains refer to Google as the main resource used in their work. The frequency with which unique digital resources are mentioned by members of each domain typically equates to a half of all resources identified. This could be attributed to the individual nature of research, which may only have one or two generic resources that deal with a particular research area, supported by a raft of highly specific resources covering the more specialised topics. The domain of classics, ancient history and archaeology reflects this view; with Google, Humbul and Perseus (very large digital libraries) offering access to information on a very broad level, but perhaps with limited depth, leaving the researcher to find more focussed information which may not be relevant to the majority of scholars within the domain.
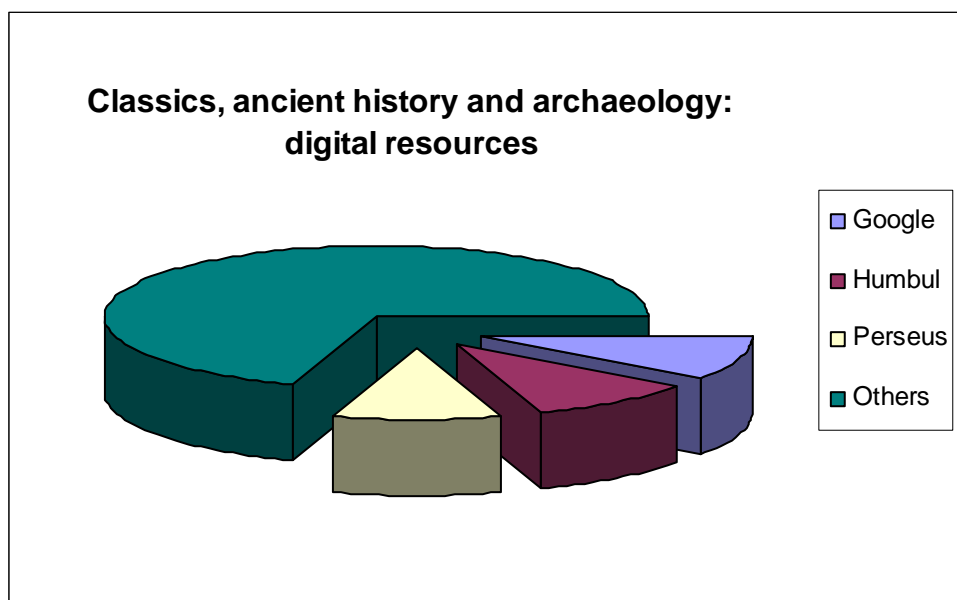
Figure 10: Digital resources identified within the domain of Classics, ancient history and archaeology.

Even within the diverse domain of philosophy, law and religion, more than half of all resources identified were only mentioned once. Although there are a greater number of generic resources within this domain, they still constitute a minority of the overall number identified.
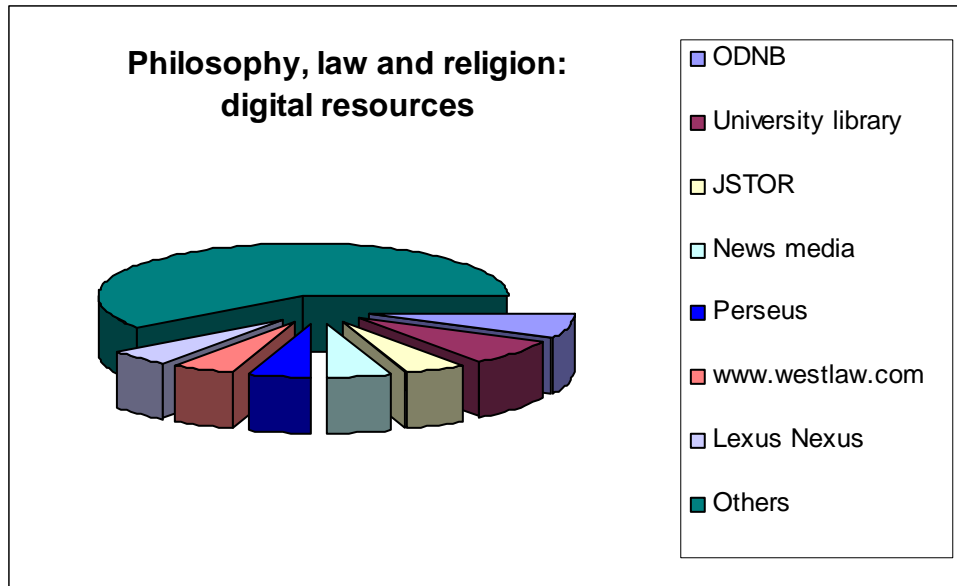
Figure 14: Digital resources identified within Philosophy, law and religion.

A full breakdown of all the resources identified by all domains is provided at **A4.7**, below. A domain specific breakdown of the resources identified is provided at **A4.8** below.

The collection and analysis of information is seen as central to the work of over 50% of the respondents, whilst data storage and archiving is not given such a high level of significance (question 23 and 24). This may be due to this facet of data retention/backup being part of the IT system currently in place at the respondent's workplace and the responsibility of an IT engineer rather than the actual individual. Further investigation is necessary to identify the reasons for this low level attribution of importance to this particular aspect of data manipulation and management.

Is data archiving once your research is complete central to your work?

22.82%   22.15%

16.11%   16.11%

12.75%

10.07%

1   2   3   4   5   6

1=Central   5=Peripheral   6=Not applicable
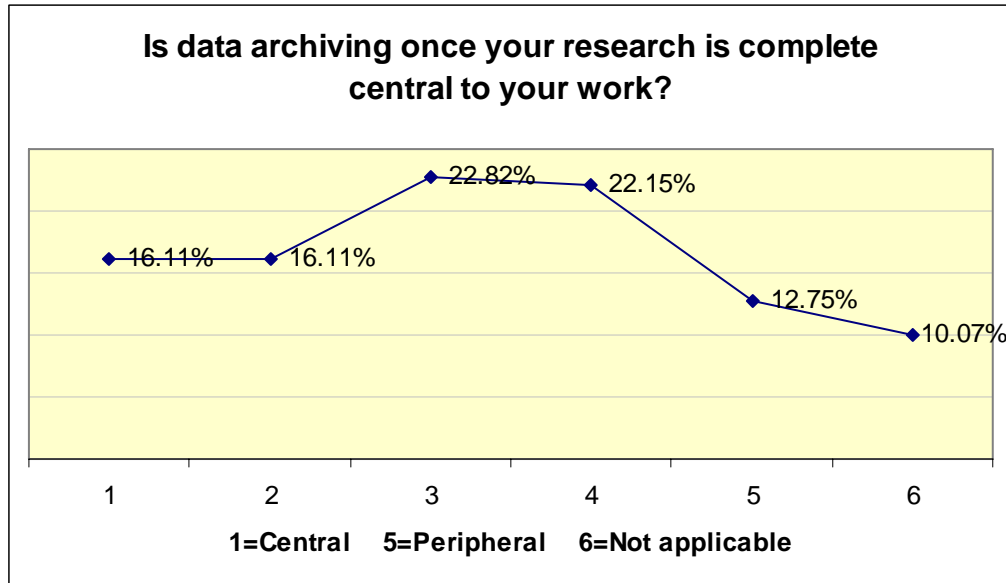
Figure 15: Responses to question 24

A4.4 Aspects of work that may be facilitated by portals

There are various important aspects of an academic's work that may be facilitated by a portal. The following section gathered data on what level of importance is attached to these areas.

General scholarship was seen as central or almost central to over 60% of respondent's work. Other academic pursuits such as debating/hypothesising, peer review, presentation of work, project collaboration and networking was also seen as more central than not to the majority of respondents.

Is general scholarship central to your work?

- 33.56%
- 32.21%
- 20.13%
- 8.72%
- 0.67%
- 4.70%

1   2   3   4   5   6

1=Central   5=Peripheral   6=Not applicable
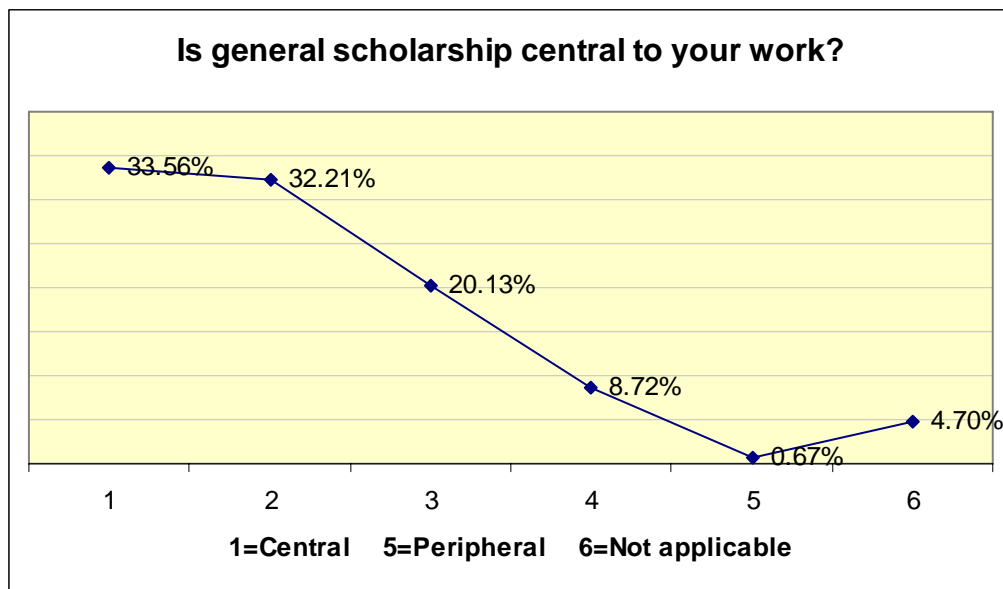
Figure 16: Responses to question 25

More administrative functions such as supervision of students and projects, direct project management, staff appointment/appraisal, writing grants, responding to tenders and consultancy work were not given the prominence of the earlier tasks or seen as not applicable.

Is staff appointment/appraisal central to your work?

41.61%

16.78%   16.78%

8.05%

6.04%

10.74%

| 1 | 2 | 3 | 4 | 5 | 6 |

1=Central   5=Peripheral   6=Not applicable
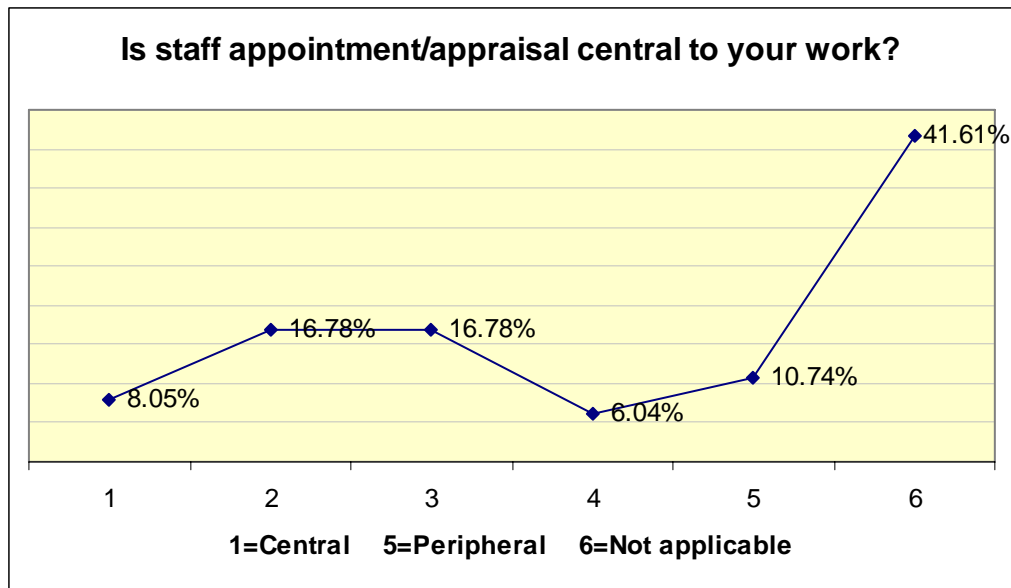
Figure 17: Responses to question 35

One comment from a respondent on this aspect of the questionnaire may hold the key "With many of these, there is no choice…" which would infer that the administrative functions are seen as a necessary chore, rather then a central facet of research activities. Other activities considered part of general scholarship and research were identified as: writing books/articles, keeping up to date with a subject area, partnership searches and thinking creatively.

A4.5 Site-specific questions

The next series of questions endeavours to ascertain more about what the respondents' wants and needs are from a website or portal, and what provision is made for them by way of resources. The data obtained from the first question provides a snapshot of how many respondents came from the AHDS websites as opposed to a variety of others. As less than 50% arrived from either AHDS or Humbul, these figures will have to be taken into account when looking at the data obtained from any following questions and making judgements as to level of provision/quality of resources; we do not have knowledge of all other originating sites. This is highlighted by free text responses in question 55 where there are a number of comments similar to "I came straight to this link via an email hyperlink so there is no site to comment on".

From what site did you enter this questionnaire?

Legend:
- AHDS
- 2. Humbul
- 3. Other

53%

29%

18%
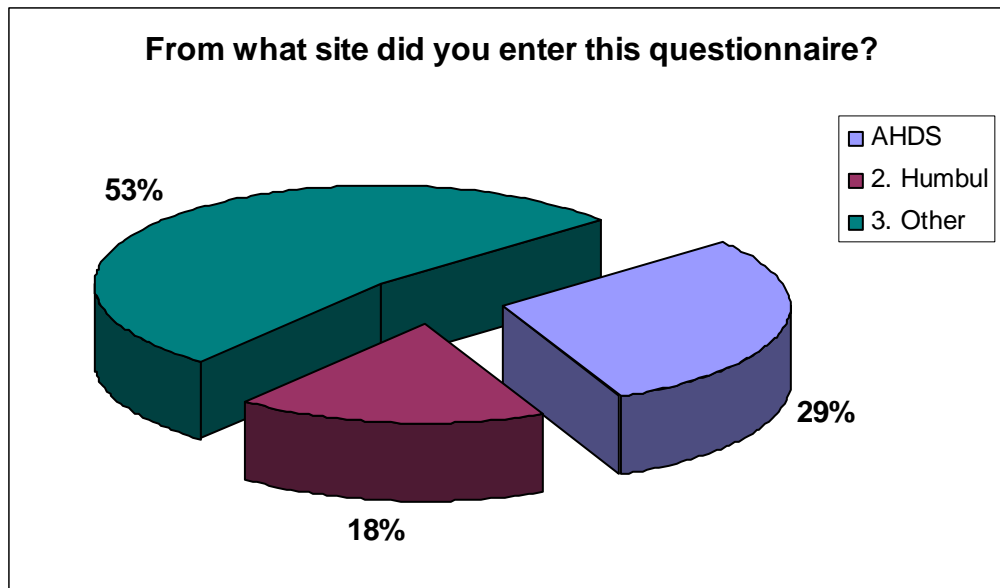
Figure 18: Number of respondents coming to the questionnaire from AHDS/Humbul

This distribution may be a product of the success of the dissemination exercise which has attracted a high number of respondents from a variety of backgrounds. The following questions concerning how the respondent found the site and how often they visited it may not relate to the AHDS/Humbul sites but provide information of a general nature. The main informant of the existence of a site appears to come from researcher-led means such as library pages or email discussion lists rather than provision by a tutor/supervisor or inclusion within course material. It also appears that the originating site is visited on a regular basis or better and would appear to be a main source of information or material for the respondents' research. The reasons for respondents' undertaking a search appears to be for non-teaching purposes, which may be triangulated to the earlier responses concerning disagreement on resources being useful for teaching or changing the way they teach (questions 11, 14 and 49).

Why are you searching for resources?

- Complete assignment
- Part of my research
- Provide teaching materials
- Part of my professional research
- General interest
- Other

13%   6%   23%   36%   9%   13%
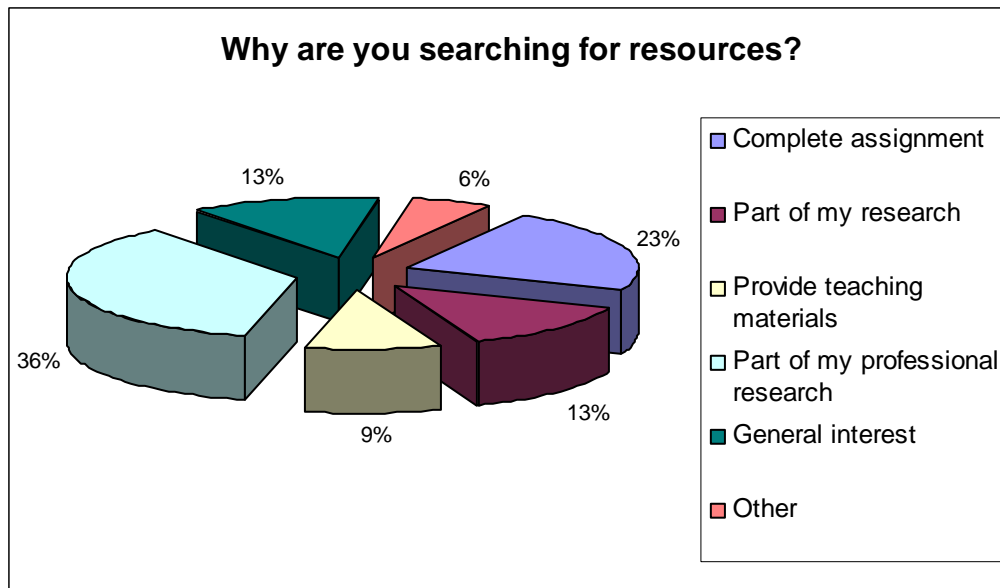
Figure 19: Respondents' reasons for conducting a search.

Although some respondents did not use, or had not seen either the Humbul or the AHDS sites, their answers still provide an insight into what is expected/required of a site dedicated to provision of resources/information for researchers.

A4.6 Conclusion

The profile of a 'normal' user is a moderate user of the web who uses digital resources extensively within their work and engages in research activities over teaching. They consider their research to be cross domain and find service provision by their University to be sufficient for their needs by using a combination of Library and VLE services supported by a scholastic centred search engine. They do not consider data storage to be a significant factor in their work, nor do they believe the administrative functions to be of primary importance. They consider their primary research websites to provide a good level of service with an adequate selection of resources which are of sufficient quality for their needs. These conclusions are confirmed by the AHDS Performing Arts online questionnaire undertaken at nearly the same time.

A4.7 List of most useful sites/digital resources

Figure 20

University/Library/OPAC/COPAC	43	ABC-CLIO	1
Google/Scholar/Images	26	Asians in media	1
JSTOR	14	ATLA	1
AHDS/Humbul	11	auditorium.ru	1
Web of knowledge	8	Bibliothèque nationale	1
News media (CNN, NYTimes)	6	Blackwell journals	1
VLE	6	Blackwell's Synergy	1
Lexus-Nexus	5	bps	1
LION	5	British History Online	1
British Library	4	British nursing index	1
National Archives	4	Business source premier	1
18thC online catalogue	3	cdwow.com	1
Amazon	3	CHILDES	1
Archives hub - mimas	3	Commonwealth war graves commission	1
BUBL	3	Connect	1
Mintel	3	Cordis	1
ODNB	3	CSA suite of databases	1
OED	3	CTHeory	1
Perseus	3	delicious	1
Web of science	3	Dept. of Health	1
wikipedia	3	digimap	1
worldcat	3	Ebay	1
www.bbc.co.uk	3	Ebrary	1
www.westlaw.com	3	ed.ac.uk	1
Yahoo	3	emerald	1
Access to archives	2	fashion-era.com	1
ACM	2	Franciscan Archive Library	1
Bodleian	2	gallica.bnf.fr	1
Early English books online	2	George Boree's site	1
EBSCO	2	Hero	1
FirstSearch	2	Historic-cities.huji.ac.il	1
Groove online	2	h-net.org	1
JISC/JISCmail	2	hnn.us	1
Library of Congress	2	IEEE library	1
Medieval sourcebook	2	IHR bibliography	1
MetaLib	2	Infotrac	1
psycINFO	2	Ingenta	1
Pub Med	2	keynote	1
RILM	2	L'année philologique	1
Science direct	2	leo	1
Sosig	2	Literature resource centre	1
Thesaurus Linguae Graecae	2	Masters of Photography	1
Voice of the shuttle	2	materials explorer	1
www.keynote.com	2	National Archives of Scotland	1

NetSERF	1	www.atpm.com/Back/atpo.shtml	1
nls.uk	1	www.bei.com	1
nursing standard	1	www.bl.uk	1
oldbaileyonline.org	1	www.bloglines.com	1
ORB	1	www.bnf.fr	1
Oxford reference online	1	www.ccel.org	1
Palatine	1	www.cilip.org.uk	1
Parip	1	www.clir.org/	1
pastmap	1	www.cnpq.br	1
PINAKES	1	www.creativecow.net	1
Powys digital library	1	www.dfes.gov.uk	1
Project muse	1	www.dfhdata.de	1
rand.org	1	www.economist.com	1
RDN virtual training	1	www.elcastello.org	1
Rhizome	1	www.erpanet.org	1
RISM	1	www.euromonitor.com	1
RNT	1	www.givemefootball.com	1
Royal Historical Society Bibliography	1	www.grovemusic.com	1
Royal Holloway 'golden pages'	1	www.history.ac.uk	1
Salidaa website	1	www.hotmail.co.uk	1
scotlandspeople.org	1	www.institutocervanctes.es	1
sparknotes	1	www.interpares.org	1
Stanford	1	www.kpmg.com	1
The Routledge Wellesley index	1	www.literatura.org	1
The Shakespeare collection	1	www.opendemocracy.com	1
Times supplement	1	www.peib.org.uk	1
UK sensitive HE MAP	1	www.pwcglobal.com	1
Victorian web literature	1	www.theatremuseum.org	1
www.3dcafe.com	1	www.thelatinlibrary	1
www.aag.org	1	www.ubu.com	1
www.abebooks.co.uk	1	www.victoriandatabase.com	1
www.affla.com	1	www.vogue.co.uk	1
www.ask.com	1	ZKM	1

A4.8 Breakdown of identified digital resources by domain.

All resources that were only mentioned once by researchers in each domain are grouped within the 'Others' category.
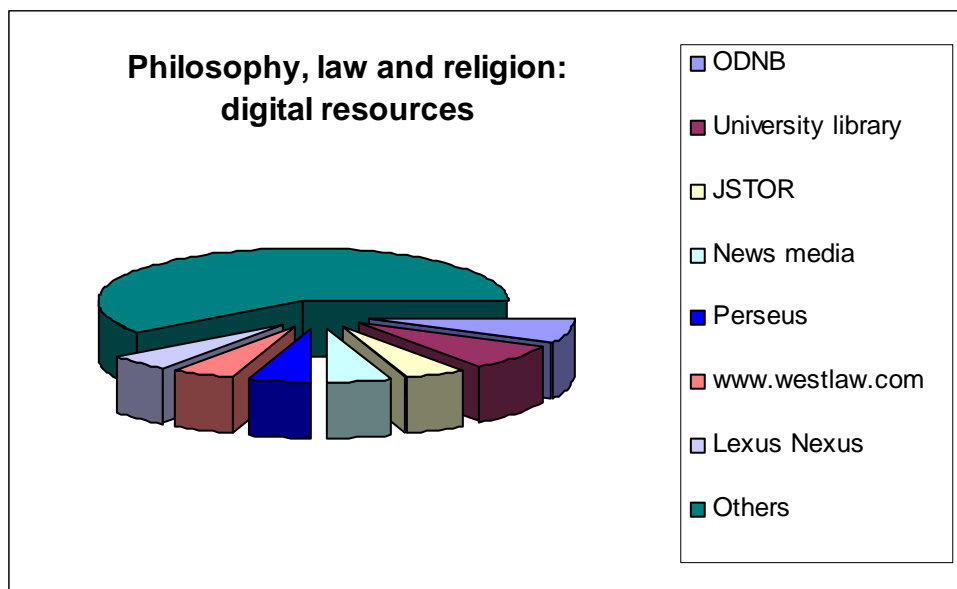
Figure 21.1 Philosophy Law and Religion Digital Resources
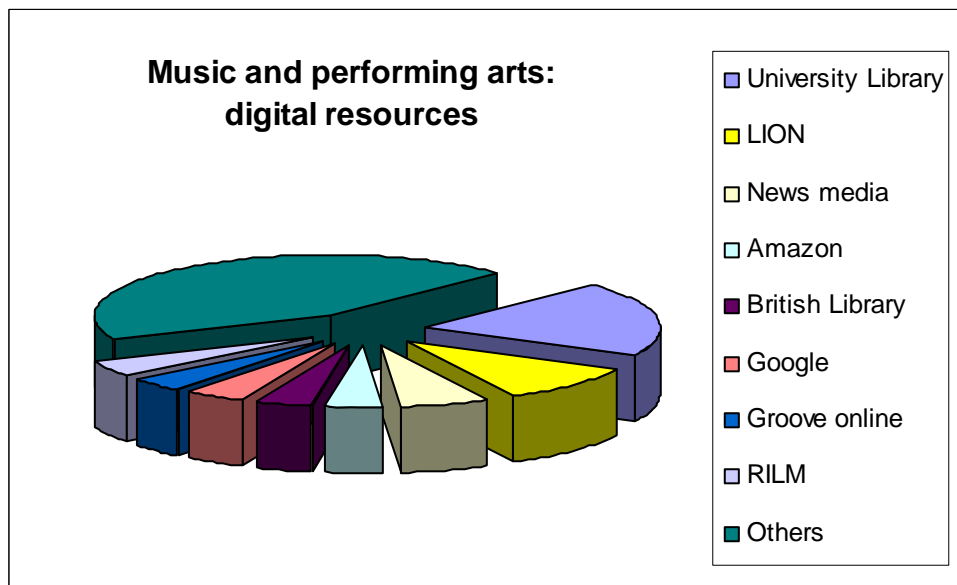
Figure 21.2 Music and Performing Arts Digital Resources

Visual arts and media: digital resources

- Google
- LION
- Universtiy Library
- ODNB
- Others
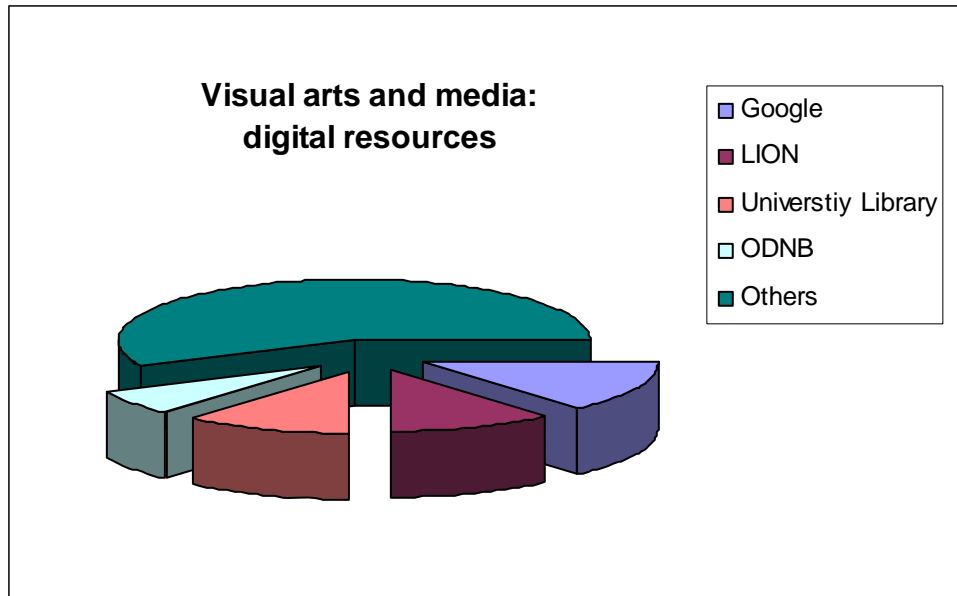
Figure 21.3 Visual Arts and Media Digital Resources

Modern languages and linguistics: digital resources

- University library
- AHDS
- Others
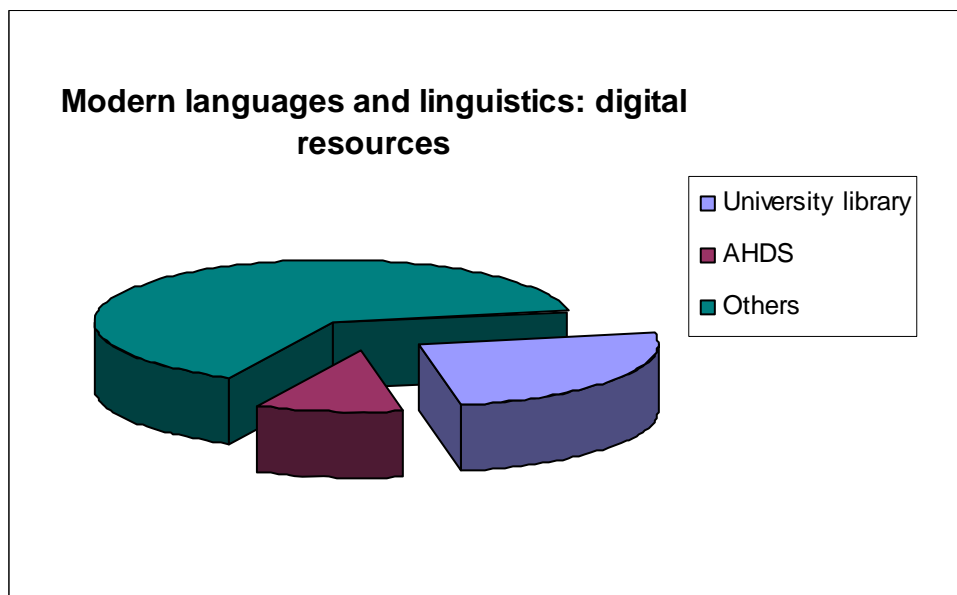
Figure 21.4 Modern Languages and Linguistics Digital Resources
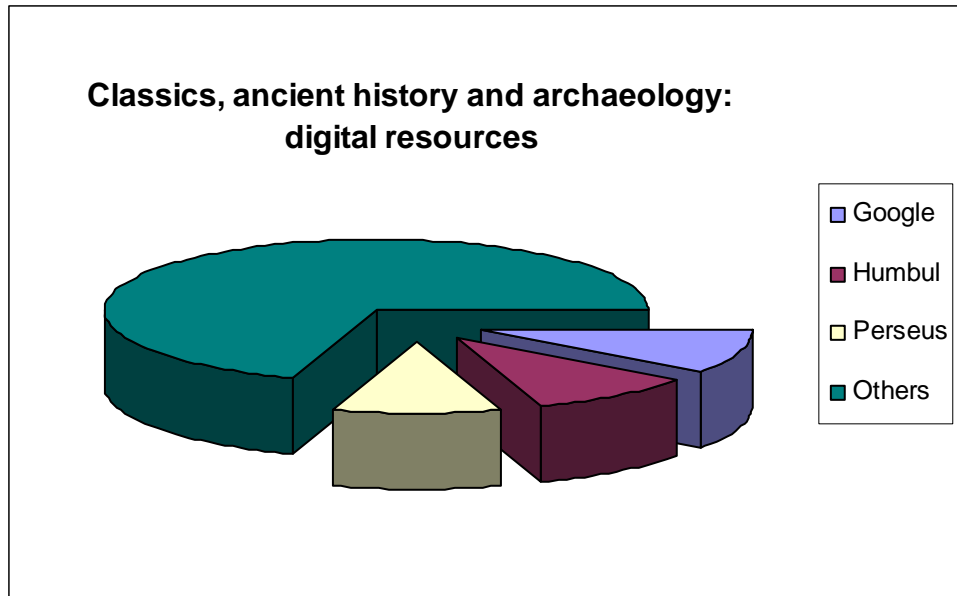
Figure 21.5 Classics Ancient History and Archaeology Digital Resources
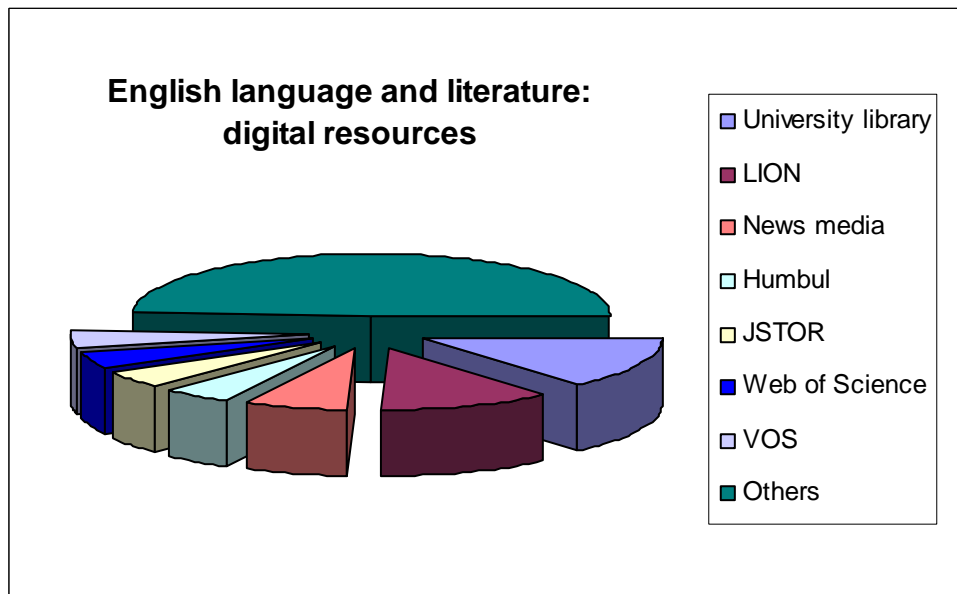
Figure 21.6 English Language and Literature Digital Resources

Librarianship, information and museum studies: digital resources

- Google
- University library
- Humbul
- JSTOR
- Others
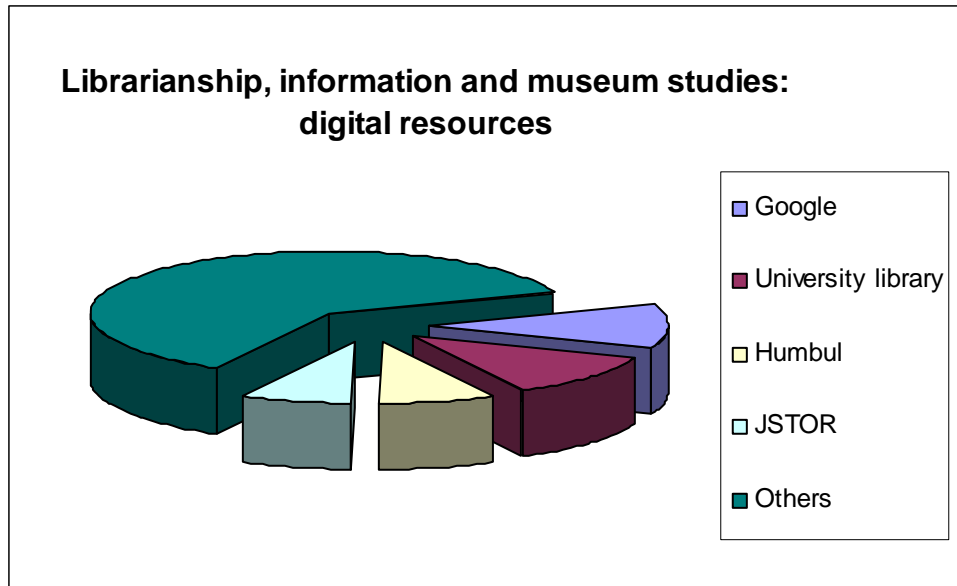
Figure 21.7 Librarianship, Information and Museum Studies Digital Resources

Medieval and modern history: digital resources

- University Library
- News media
- Google
- ODNB
- Archives hub
- Bodleian
- EEBO
- Humbul
- Others
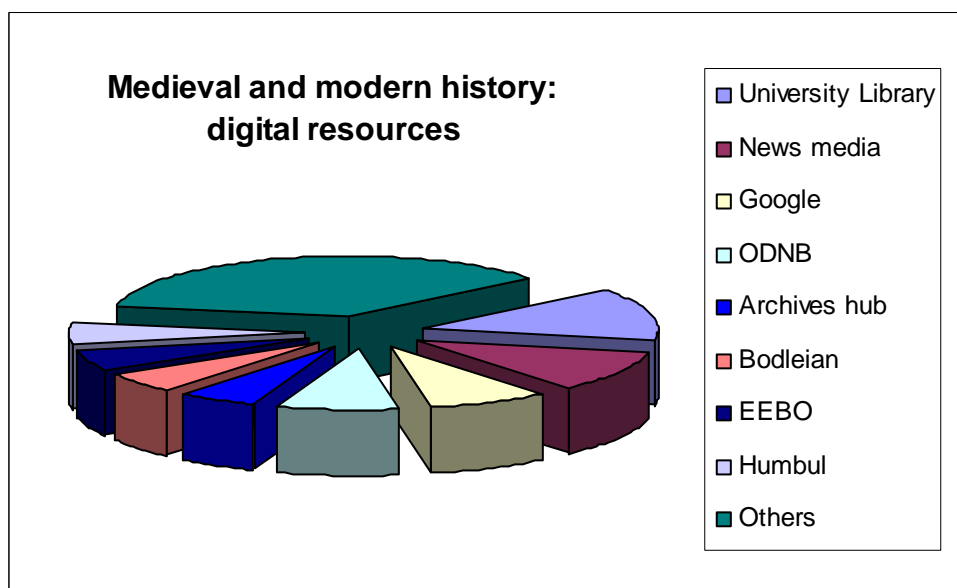
Figure 21.8 Medieval and Modern History Digital Resources

A4.9 Research Portals in the Arts and Humanities Questionnaire

"How does the arts and humanities research community find and exploit the internet resources it needs? This survey will be open from 1/12/05 to 30/4/06. Technical developments now make it possible to refine, personalise, cross link and render interactive online information gateways. We want to examine current user information search/access strategies and patterns and develop demonstrators of interactive gateways to investigate future user requirements for advanced information services that will serve to facilitate greater take and up use of these resources. Recommendations will be made to the AHRC on future policy development. This is your chance to influence how the work proceeds, so please could you answer every question (even if it is a 'not applicable') as this will help us with data analysis. There is also a prize draw with a chance to win £100. Details of how to enter are at the end of the questionnaire. Please click the 'Next' button to begin the questionnaire."

Section 1: About you.

1: What kind of researcher are you? (Circle One)
- Independent.
- A post-doctoral researcher at an HE institution.
- A lecturer at an HE institution.
- An academic-related support worker at an HE institution.
- A self-directed postgraduate student.
- A taught postgraduate student.
- Other (free text)

2: What age are you?
- 18-21
- 22-30
- 30-40
- 40-50
- 50-60
- 60-65
- 65 and above

3: What is your gender?
- Male/Female

4: Are you based in the United Kingdom?
    YES/NO (If no- where are you based? - free text)

5: How often do you use the web during the working week?
- Every day- four hours or more
- Every day- less that four hours
- Several times a week

6: How long have you been using the web?
- Over last 10 years
- Last 10 years
- Last 5 years
- Within the last year

7: Your research domain(s):

Please tell us what domains the research you undertake falls within. Please also prioritise the domains you work within by selecting a number from the related scale: 1 being HIGHEST priority and 8 being LOWEST priority

Classics, ancient history and archaeology	Priority: select 1-8	Modern languages and linguistics	Priority: select 1-8
Visual arts and media	Priority: select 1-8	Librarianship, information and museum studies	Priority: select 1-8
English language and literature	Priority: select 1-8	Music and performing arts	Priority: select 1-8
Medieval and modern history	Priority: select 1-8	Philosophy, law and religious studies	Priority: select 1-8

Section 2: Use of digital resources.

Please tell us how extensively you use digital resources in your studies, teaching or research. Please indicate how far you agree with the following statements (1 indicates strongly agree, 5 indicates strongly disagree)

8: I could not do my academic work without access to digital resources.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

9: I use computational techniques or tools extensively in my academic work.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

10: I use digital resources extensively in my academic work.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

11: Digital resources are useful for teaching but not for research.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

12: Digital resources are not of a sufficiently high quality to be useful to me.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

13: Digital resources have changed the way that I do my research.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

14: Digital resources have changed the way that I teach.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

15: The internet has made it easier to gain access to scholarly resources.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

16: I use the internet to find textual material, I print it out, then read it.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

17: I find interactive digital content most useful for my work.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

18: I have to use too many digital resources and would prefer more printed material.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)
19: Please tell us which three websites or digital resources you have found most useful in your academic work.
(Free text)
20: Any other comments?
(Free text)
Section 3: Research activities.
For each of the activities listed below please indicate how much they are central/peripheral to your research.
21: Working with data and information. (Central 1 2 3 4 5 Peripheral)
(Central 1 2 3 4 5 Peripheral **N/A**)
22: Data and information analysis
(Central 1 2 3 4 5 Peripheral **N/A**)
23: Data storage – while research undertaken.
(Central 1 2 3 4 5 Peripheral **N/A**)
24: Data archiving – once research phase completed.
(Central 1 2 3 4 5 Peripheral **N/A**)
25: General scholarship (i.e. keeping abreast with thinking in your field(s) as opposed to looking for specific information for a project).
(Central 1 2 3 4 5 Peripheral **N/A**)
26: Working with others.
(Central 1 2 3 4 5 Peripheral **N/A**)
27: Informal networking, at conferences, workshops, within research communities.
(Central 1 2 3 4 5 Peripheral **N/A**)
28: Debating, hypothesizing- the communal development of ideas.
(Central 1 2 3 4 5 Peripheral **N/A**)

Section 4: Dissemination / publishing.

29: Presentation of work in progress, perhaps to steering committees, colleagues, etc, or maybe for peer criticism, performance rehearsals, etc.

(Central 1 2 3 4 5 Peripheral **N/A**)
30: Formal publication / presentation of work, perhaps at conferences, within journals, at a formal exhibition, to an audience made up of the general public etc.
(Central 1 2 3 4 5 Peripheral **N/A**)
Section 5: Peer review, supervisory and managerial activities.
31: Peer review and criticism of books, articles, performances, exhibitions, etc.
(Central 1 2 3 4 5 Peripheral **N/A**)
32: Supervision of students.
(Central 1 2 3 4 5 Peripheral **N/A**)
33: Supervision of projects- advising steering groups, project boards etc.

(Central 1 2 3 4 5 Peripheral **N/A**)

34: Direct project management- e.g.: resource allocation, scheduling, budgeting etc for ongoing research projects.

(Central 1 2 3 4 5 Peripheral **N/A**)

35: Staff appointment / appraisal: either as an invited interview panel member or recruiting and managing staff in your own research group.

(Central 1 2 3 4 5 Peripheral **N/A**)

Section 6: Grants, tenders and consultancy.

36: Writing grant applications.

(Central 1 2 3 4 5 Peripheral **N/A**)

37: Responding to Invitations To Tender or offers of work in the public / private sectors.

(Central 1 2 3 4 5 Peripheral **N/A**)

38: Undertaking consultancy work for external clients.

(Central 1 2 3 4 5 Peripheral **N/A**)

Section 7: Have we missed anything?

39: Please outline any other activities that you would consider part of your research work in the box below.

(Free text)

If you have come to this questionnaire from either AHDS or Humbul please answer the following section. If you came to this questionnaire through any other route, please go to question 57

Section 8: AHDS and Humbul

40: This questionnaire is linked to a number of sites, but we need to know about the one from which you have just come. Can you please tell us what site it was.

- AHDS
- Humbul
- Other (Free text)

41: How did you find out about that site (the one you have just come from: AHDS, Humbul, etc)? Please tick as many boxes as apply.

- Recommendation by course tutor (please specify course in the box below)
- Recommendation by PhD supervisor
- Library webpage
- Departmental web page
- Referral by a friend or colleague
- Recommendation by computing support officer
- Email discussion list - please specify in box below
- Printed promotional material
- Other - please specify in the box below

(Free text)

42: How often have you visited that site?

- First visit
- Rarely
- Regularly (several times a year)

- Frequently (every month)
- More often - please specify

(Free text)

43: Why are you searching for resources?

- To complete an essay or assignment
- As part of my dissertation research
- To provide teaching materials
- As part of my professional research
- General interest
- Other (please specify)

(Free text)

Section 9: Please tell us your views on the usefulness of the site. Please indicate how far you agree with the following statements (1 indicates you strongly agree, 5 that you strongly disagree).

44: The site is a very helpful way to find the resources I need.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

45: The site covers a good range of academic disciplines

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

46: The site lists too many resources.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

47: Resources I can find through the site are not sufficiently specialised.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

48: The range of resources I can find through the site is limited.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

49: Resources that I have found through the site are helpful for my teaching.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

50: Resources I have found through the site are helpful in my research.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

51: I will return to this site when I need to find other resources.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

52: I can find appropriate resources for my specialist subject.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

53: I have found resources through the site that I would not otherwise have known about.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

54: The site is easy to use and navigate.

(Strongly agree 1 2 3 4 5 Strongly disagree **N/A**)

55: Please give any other comments here.

(Free text)

Section 10

56: Would you like to enter the prize draw?

- Yes
- No

57: Would you be willing to be interviewed about your views?
- Yes
- No

58: Would you like a copy of the final report?
- Yes
- No

If you have answered yes to any of the previous 3 questions, please supply your email address.
(Free text)

Appendix A5 Work-Package 2: Web-Log Analysis Report

WP2 Report prepared by Robb Ross

A5.1 Introduction

This report is prepared on the basis of web-log data provided by the AHDS and RDN subject portals over the following periods:

AHDS Central Server Access Logs [February-September 2005]
AHDS History Server Logs [*data still under analysis*]
AHDS Visual Arts Logs [*data still under analysis*]
Humbul Server Logs [January-December 2005]
Artifact Server Logs [October-December 2005]

The data provided is difficult to analyse, and for a number of reasons. Because of the distributed nature of the AHDS service, many users access their online resources through the particular service rather than the central server. In the case of AHDS History, however, its server logs are rolled up with the Data Archive. They were able to strip out for us the traffic that was not relevant to our needs. In the case of AHDS Archaeology, where the traffic is heavily influenced by non-HEI needs, we did not feel that the analysis would be relevant. In all the AHDS data, there is internal traffic between the AHDS sites that we have not been able to strip out from our analysis. In the case of Artifact, the server logs were not archived prior to October 2005 and so we have only a fragment of the picture to go on. No web-log data was forthcoming from AHDS Language, Literature and Linguistics, or from AHDS Performing Arts. Although the JISC requires some statistics from web-log activity to be published from the services that it supports, they are not published in a coherent fashion. We would expect to be able to recover sample statistics of the following from the Annual Reports of the services as 'surface-measures' of user traffic:

- o   Site Visits
- o   Total Page/\Item Views per month
- o   Average No of Pages/Items consulted per day
- o   A statistical reflection of particular function-usage. In the case of AHDS this may be collection downloads. In the case of Intute, it may be registered users.

In reality this is not the case. The statistical analysis of these data-logs was undertaken for the project by Dr Paul Huntington of CIBER, UCL. Based on that evidence, this Report has been prepared by the Project, which is responsible for its conclusions.

A5.2 Web-Log Analysis Methodology

Web-server logs record simple traffic statistics and data such as number of page requests per month and originating addresses of page requests. Deep-log analysis (DLA) uses web logs from a server and, following a normal process of analysis, links the information with

site-user profiles, or demographics, to produce a 'deeper, more meaningful data' picture of overall site usage. It is a four stage process:

1)   Data definition, recording procedure and statistical significance are agreed.
2)   A series of pre-defined metrics are used to ensure the data is analysed in line with organisational goals and policies.
3)   Enrichment of usage data with demographic data.
4)   Identification of questions concerning information seeking behaviour that need to be asked by questionnaire, interview or observation.

The **working metric definitions** used in this report are:

User: A user is effectively a computer; sometimes that computer represents an individual, (i.e. a professor in his office), in other cases a number of people (i.e. students in the library). User identification can be based on a combination of "IP" number and browser details or by use of cookies.

Sessions: They are identified in the logs by a session identification number. Logs include a session beginning tag and a session ending tag, which enables time calculations as well. Items viewed/requests made. The key usage sub-metrics are: type of items viewed, number of items viewed in a session and return visits. These sub-metrics offer extremely good platforms for characterising and comparing the information seeking behaviour of sub-groups of users because generalisations based upon millions of users, while sounding impressive, can prove very misleading, camouflaging possibly big differences between individual user groups, like that between students and professors. A complete item might be all the pages, charts etc. from an article, and this is recorded as a single item and hence the digital library logs are quite different from traditional server log files that record pictures and text documents separately. The logs may also recorded views to the home page and a returned search screen.

Items viewed/requests made: This is defined as a 'complete' item returned by the server to the client in response to a user action. Typically this might be a menu page or a search screen. Logs do not record all items viewed by the client since, once seen, the item will be cached to the clients' machine. If a client returns to view the page that view will be made from the copy in the cache and not from the server. A page will remain in the cache for a variable length of time.

Robots: Web-logs often record 'robot' users to sites. These are mechanical agents, used mainly be search engines, to index web-pages. Robots should report to the site's 'robot.txt' document, which identifies the accesses by this IP number as a robot and informs the robot as to what pages to index. Several robots were identified in the course of the CIBER analysis which did not conform to that convention and these were also stripped out. Views to automatic feeds were also stripped out from the analysis.

Internet protocol (IP) numbers: These are identities that facilitate an access to view items on the internet. IP numbers also act like registration numbers and can be used to access additional information about the user in a process called reverse DNS (Domain Name Server) lookup. This process, when successful, reveals the user's organisation name, the type and the location of organisation. However many users mis-register. So, for example, a UK user may register for a US-style domain name or a net-provider will often register as a commercial organisation. Further, not all IP numbers can be identified by this process. Though these difficulties limit, they do not negate the usefulness of such data. Academic institutions, in particular, rarely mis-register either their location of organisational grouping.

Referrer Links: These are the identified site link from which the user accessed the site being investigated.

A more powerful way of examining the number of items viewed is to categorise search sessions by the **number of items viewed**. This is called 'site penetration'. Research on the subject has shown that many web users graze lightly, examining just a few items/pages before they leave with no substantial content consumed, although knowledge might have been gained [Nicholas et al, 2004c]. High levels of penetration can be assumed when there is evidence of:

  a) 'natural movement' through the site
  b) a massive choice of data on offer
  c) the investigative nature of some information-seeking
  d) the presence of an embedded search engine and other retrieval aids.

Returning to a site also constitutes evidence of conscious and direct use. However, research on that subject suggests that people view only a small proposition of a site's contents and, further, return to it very rarely [Nicholas et al, 2004c]. In theory, how frequently they return should depend on the nature of the site – a newspaper site, for instance, might be expected to obtain more return visits. But there is no natural frequency for any particular kind of site. But, in the case of academic information-seeking behaviour, one might expect a more developed repeat-behaviour (in order to satisfy reiterated information needs) than other internet information-seekers. In general, the ability to generate useful information via DLA relies on adding user demographic data (e.g. occupation, subject specialism), via data obtained from a subscriber database (preferable) or online questionnaires (less preferable, since user data cannot be mapped so closely onto usage data). Of course, logs and user databases enable us to map the digital environment more accurately but provide little by way of explanation, satisfaction and impacts.

A5.3 Humbul Web-Log Analysis

A5.3.1 Items Viewed:

Figure 1 looks at the daily number of items viewed over 2005. Usage generally fell within the range of about 2,000 to 4,000 daily views at weekends and between 6,000 to

8,000 item views on weekdays. The year started relatively low between 3,000 to 6,000 daily views and then rose to a peak in early to mid February of 8,000 (weekdays) before declining and reaching a relatively low figure of 6,000 (weekdays) in late march early April. Use recovered over the next month before entering a decline over May to early September. Over this period the range between weekday and weekend use declined. Use picked up fairly strongly from mid-September to December.

Figure 1: Daily number of items viewed 2005.
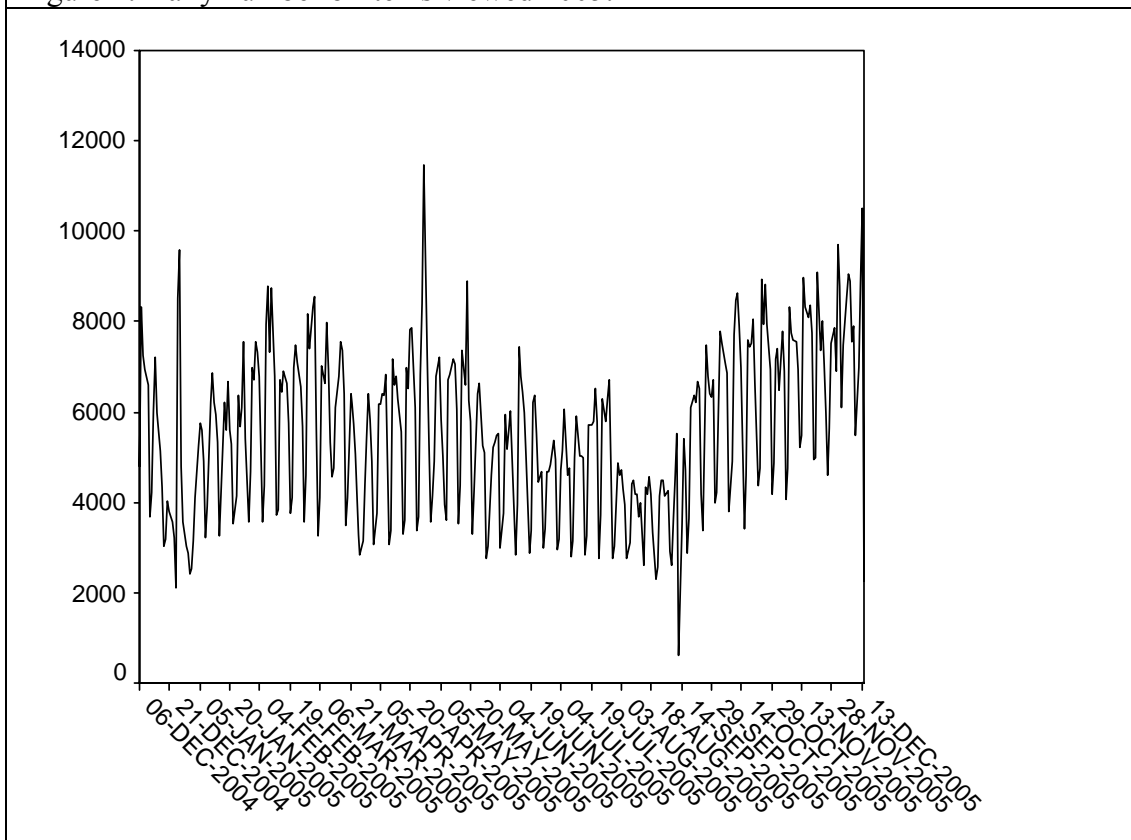
Figure 2 gives the percentage distribution of usage over day of week. Usage at the weekend was about two thirds of usage on weekdays.

Figure 2 The percentage distribution of usage over day of week
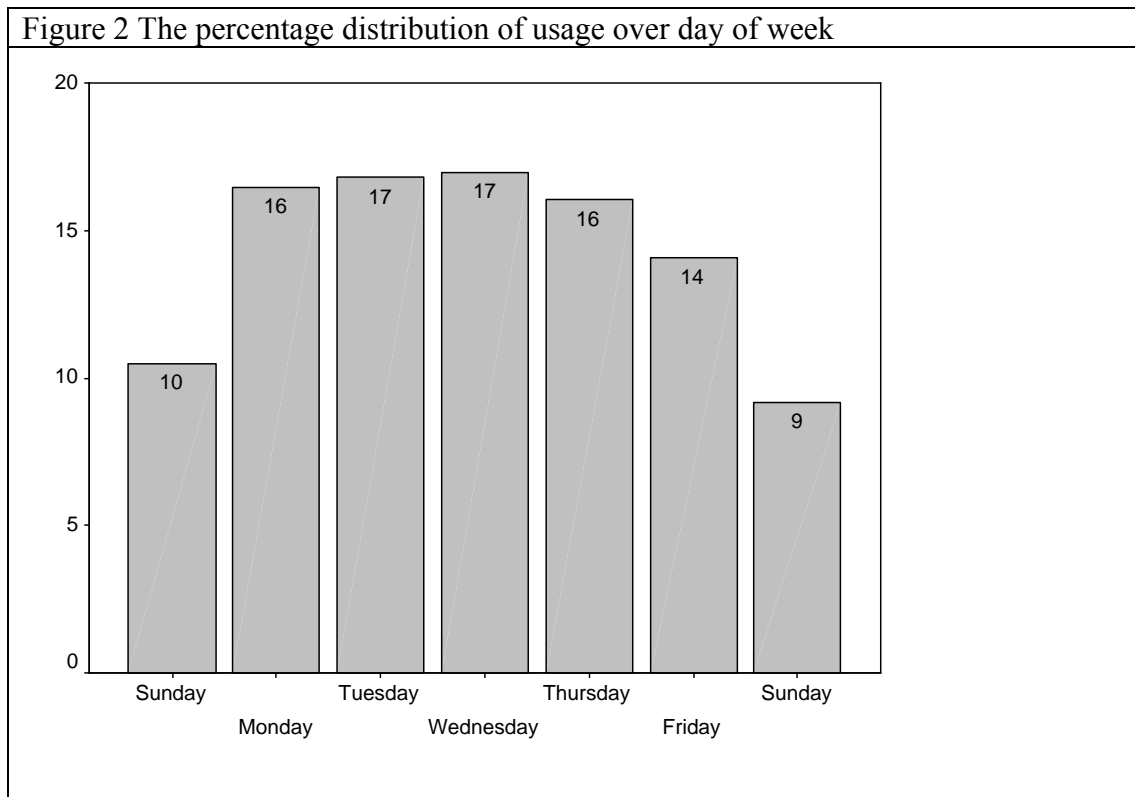
Figure 3 gives the share of usage broken down by organisation type, as retrieved from the reverse DNS lookup. About a third (32.8%) of usage is attributed to academic users, just over a third (37.3%) to commercial users and over a quarter (28.9%) of usage is attributed to net providers.

Figure 3 The share of usage broken down by organisation (DNS) type

Academic
32.8%

Non profit
.6%

Government
.4%

Net provider
28.9%

Commercial
37.3%
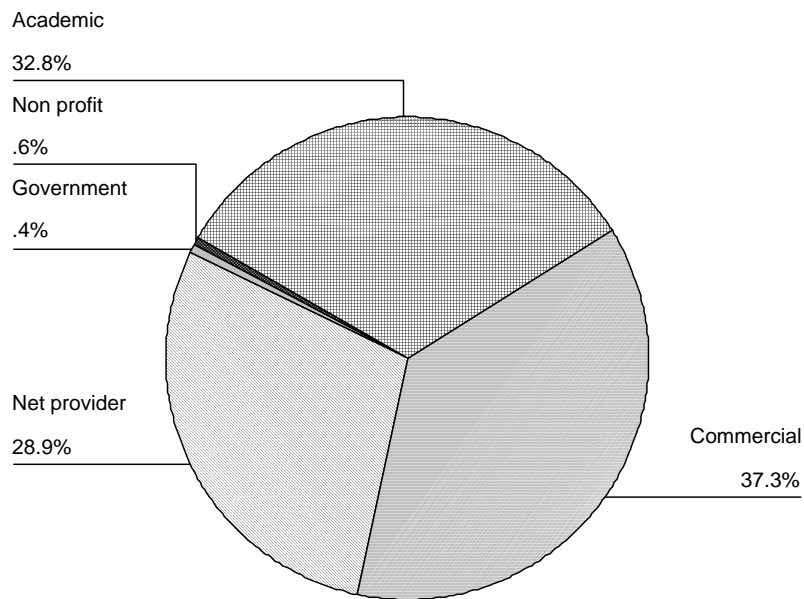
The main two countries from which users were using the Humbul services by reverse DNS look up was the US (50.1%) and UK (29.9%). The group other includes all countries that each made up less than 2% of usage.

Figure 4 The share of usage broken down by user (DNS) country code

Other
20.0%

United States
50.1%

United Kingdom
29.9%
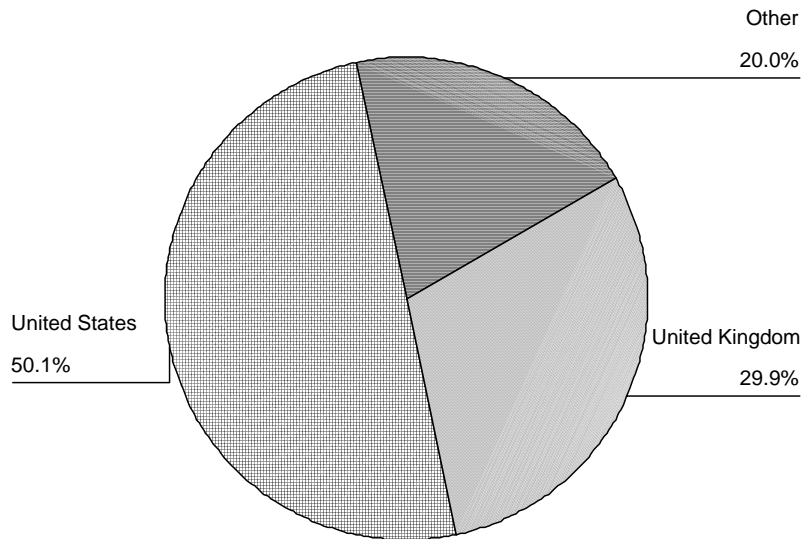
Other all countries accounting for less than 2% of use

Figure 5 gives the same information but breaks it down according to the country location by world regions. Western Europe (excluding UK) made up about 7 to 8% of usage while Eastern Europe made up between 3 to 4%.

Figure 5 The share of usage broken down by user (DNS) country codes grouped into world regions.

US

50.1%

W Europe

7.6%

Other

1.7%

Asia

2.7%

Australia/Oceania

2.3%

E Europe

3.6%

N-America

2.2%

UK

29.9%

Other: grouped regions that individually account for less than 1% of usage.
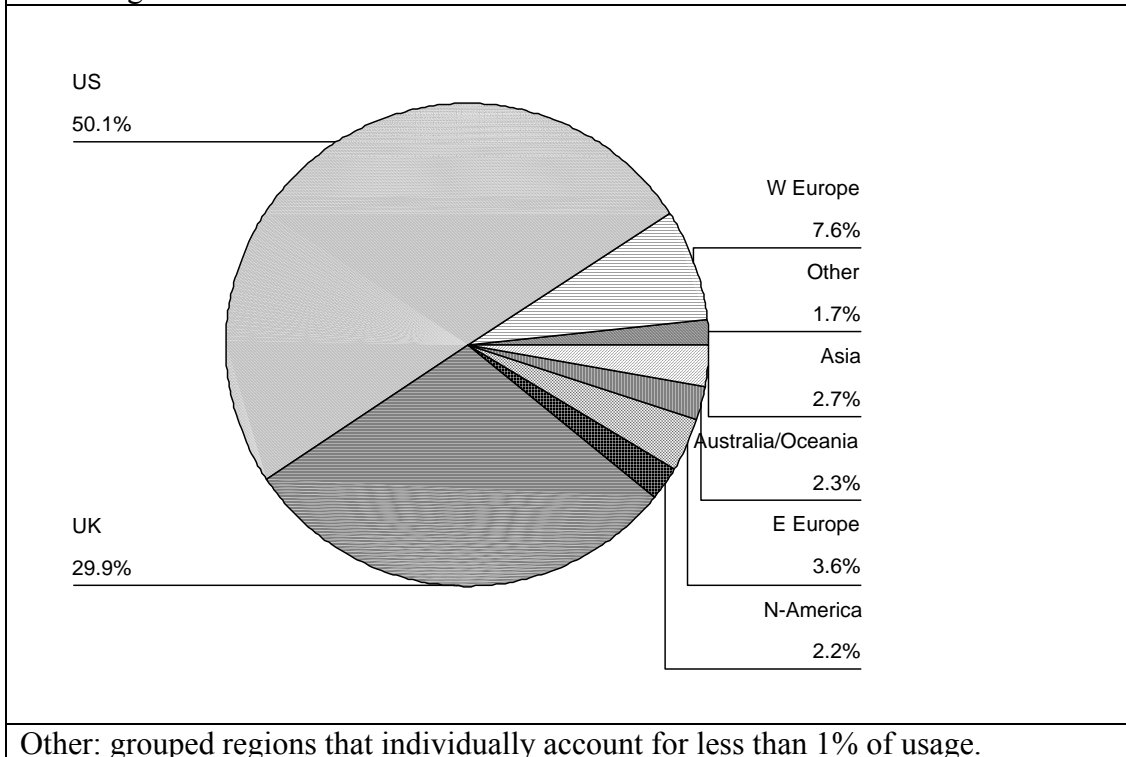
Figure 6 gives an idea of item-type viewed. The homepage, that accounted for about 11% of views, is the opening Humbul page that is viewed on opening the site at www.Humbul.ac.uk. This page includes a variety of subject links. It is defined here as the 'menu1' page. There were approximately 5.5% views to this page. This page offers links to sub-categories within the subject. Clicking on any of these links takes the user to a sub-menu (menu2) page – which made up 5.6% of views. The menu2 page offers users a list of resources to link to. Under each resource is a reduced summary, a link to the extended summary and a link to the resource. Should the user opt for the extended summary the user is taken to the ID (extended summary). This gives an extended summary of the resource and a link to the resource. About a quarter (24.6%) of items viewed were to the extended summary. About 11.5% of users activated the link to the external URI (universal resource indicator). Rather than use the menus, users may alternatively activate the on-site search facility. About 9.3% of usage related to items where the word-search appeared. Other identified items were to do with 'jobsearch' and other items that appear on the left hand menu of the Humbul homepage. Other unidentified pages made up 14% of usage. Most (about 75%) of the other unidentified group was accounted for by the following item names: describe (19%), user (14%), vts (9%), about (8%), help (8%), submit (8%), topics (7%), output (5%).

Figure 6: Distribution of item type viewed

Other not identified
14.4%

Other identified
4.2%

Menu1
5.5%

Homepage
11.1%

Search
9.3%

ID extended summary
24.6%

URI link used
11.5%

sub (menu2)
5.6%

No
2.8%

404
11.1%
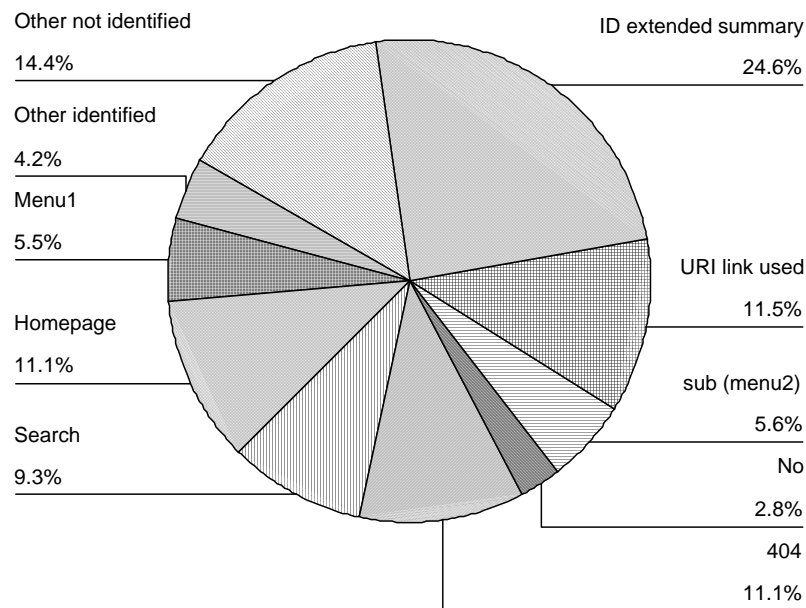
Other identified groups items where each item accounted for less than 2% of usage.

In terms of menu1 usage the following gives an idea of subject-usage. History is the most popular subject and about a quarter (27.1%) of subject-use relates to this. Other popular subjects are English (16.9%), Religion (6.5%), Humanities_a (6.2%) and Philosophy (5.1%).

Figure 6: Distribution of subject item (Menu1) viewed

History
27.1%

English
16.9%

Other
15.5%

Religion
6.5%

French
4.2%

Humanities_a
6.2%

American
4.2%

Philosophy
5.1%

Archaeology
4.5%

Classics
5.0%

Linguistics
4.7%
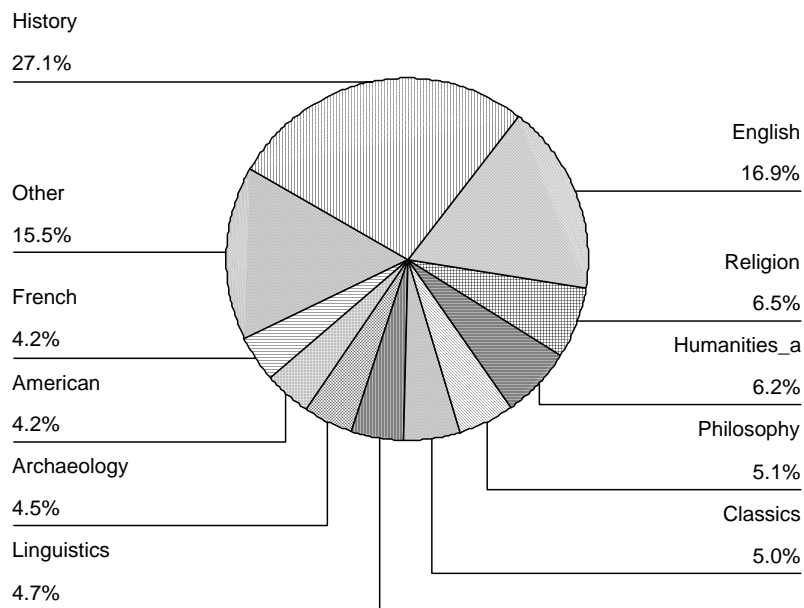
Other groups subjects that each account for less than 3% of use.

The Humbul logs also give the site address and directory of the linking resource. If the user decides to visit a resource, the logs record the site visited. About 11.5% of items viewed were users who then actively clicked through to the resource. Throughout the year, 7,463 separate resources were accessed via the Humbul site. The following Table lists the top 40 and the accompanying spreadsheet gives the full list.

Figure 7: Top 40 resource sites accessed via Humbul

URI Site	Number	Percentage
www.bbc.co.uk	4166	1.5
www.wsu.edu	2473	.9
www.geocities.com	1969	.7
www.nd.edu	1517	.6
ads.ahds.ac.uk	1216	.4
www.bl.uk	1047	.4
www.arts.ed.ac.uk	1042	.4
www.pbs.org	1031	.4
www.emule.com	936	.3
memory.loc.gov	836	.3
www.fordham.edu	813	.3
www.shef.ac.uk	811	.3
www.channel4.com	789	.3
www.newadvent.org	713	.3
www.llgc.org.uk	680	.3
www.spartacus.school	659	.2
www.luminarium.org	659	.2
etext.lib.virginia.e	649	.2
uk.cambridge.org	643	.2
www.ucl.ac.uk	636	.2
www.iwm.org.uk	624	.2
www.loc.gov	614	.2
ccat.sas.upenn.edu	606	.2
www.gre.ac.uk	599	.2
www.archives.gov.on.	575	.2
www3.oup.co.uk	573	.2
www.archives.gov	563	.2
www.accd.edu	560	.2
www.nationalarchives	559	.2
www.georgetown.edu	546	.2
www.hti.umich.edu	540	.2
www.sas.ac.uk	536	.2
www.kb.nl	520	.2
etext.virginia.edu	506	.2
www.bu.edu	504	.2
www.stoa.org	503	.2
history.hanover.edu	499	.2
raven.cc.ku.edu	490	.2
learningcurve.pro.go	485	.2
www.17thc.us	479	.2
		12.6%

In terms of referrer link, about 40% of use related to users coming in via Yahoo, 20% via Google. Other sites include Wanadoo (3.2), ox (4.9), RDN (4%), Altavista (1.8) and the BBC (1.4). There is a specific reason for the apparently disproportionate number of user coming to Humbul via Yahoo. We understand that Humbul exposed its metadata via OAI to Yahoo for them to index their aggregated collection of harvested metadata. As a result Humbul's metadata records are high in Yahoos rankings. Yahoo is a commonly used as commercial search engine of choice, particularly among non-academics in North America. So far as we are aware this is the only example of OAI metadata being made available for harvesting by the commercial search engines from the service providers. Its significant impact upon usage patterns should be noted.

A5.3.2 Part 2 Session level analysis

The following relates to the number of sessions. The site attracted between 1,500 (weekend) to 2,500 (weekday) sessions a day. The pattern of session over the year followed the same pattern as for items viewed.

Figure 8: Daily number of sessions - 2005
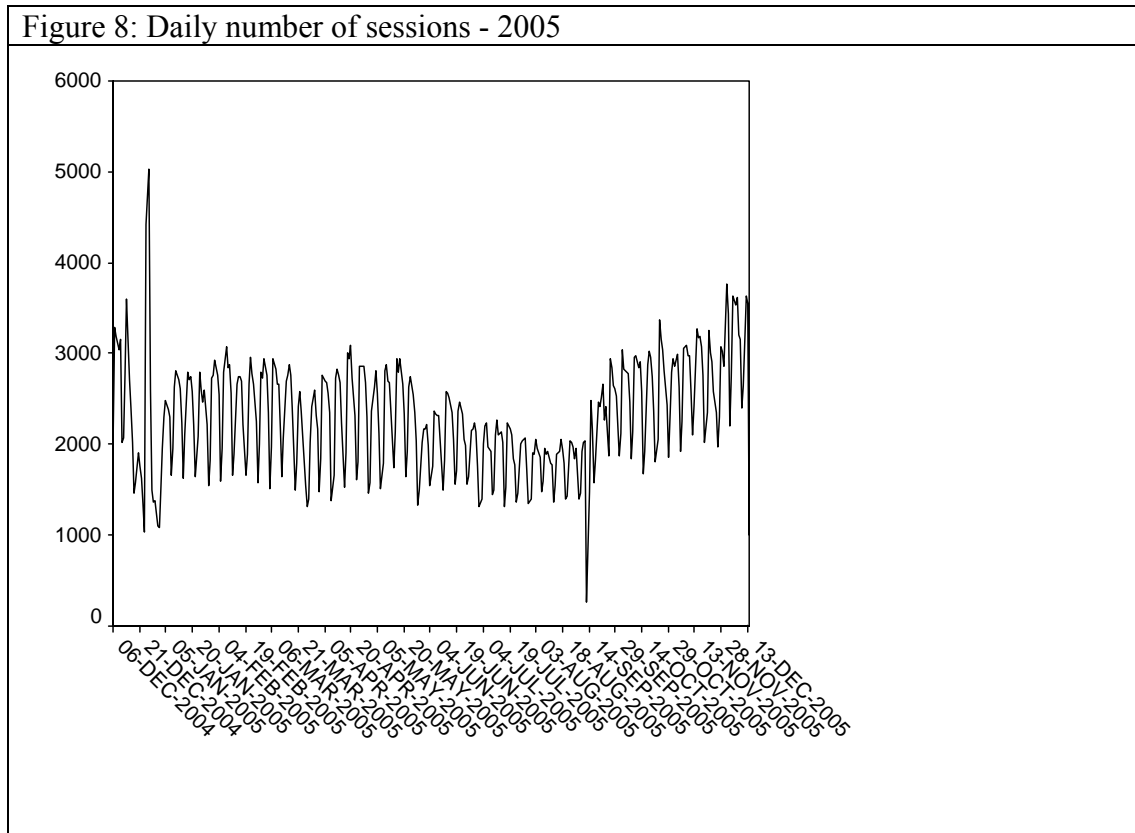
Figure 8 gives the same distribution but as a percentage.

Figure 8: Daily number of sessions - 2005
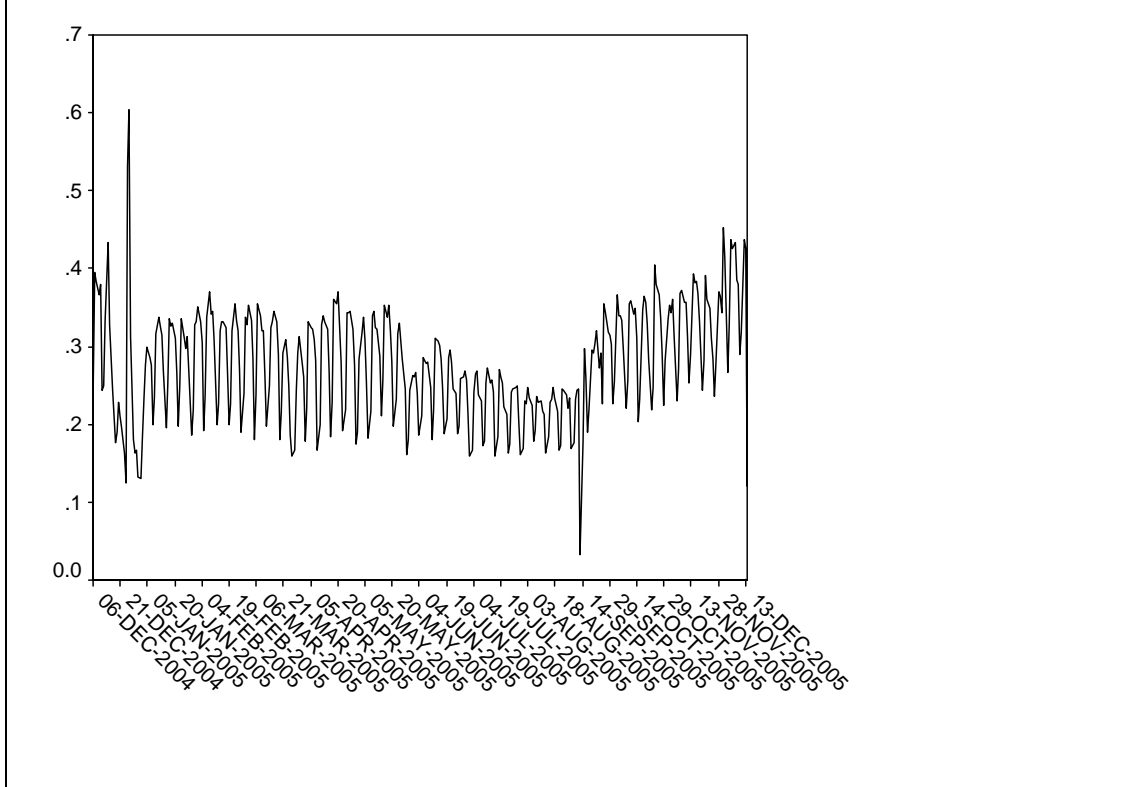
Figure 9 gives the number of sessions for each month for 2005.

Figure 9 The number of sessions for each month for 2005.
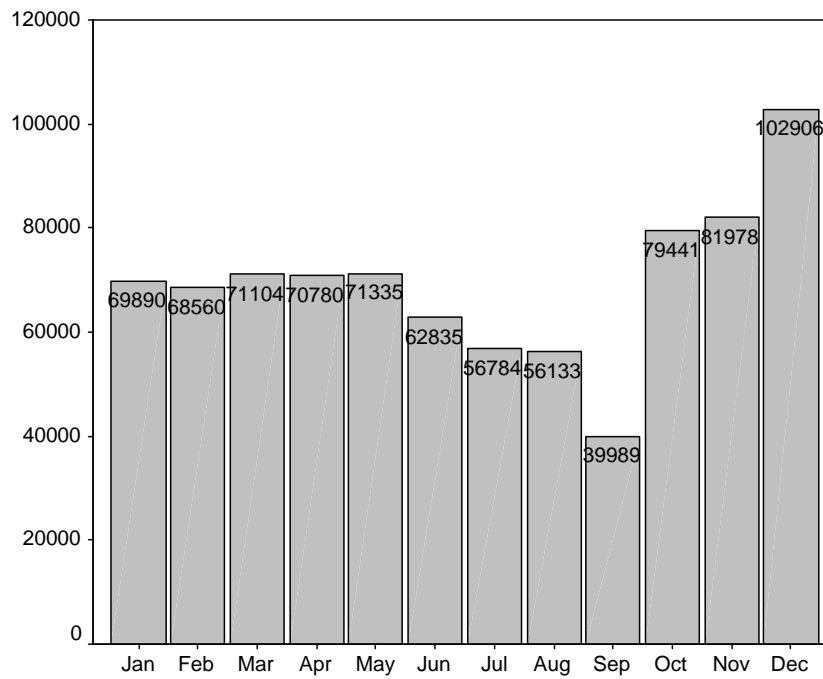
Figure 10: location of user as given by DNS registration details.
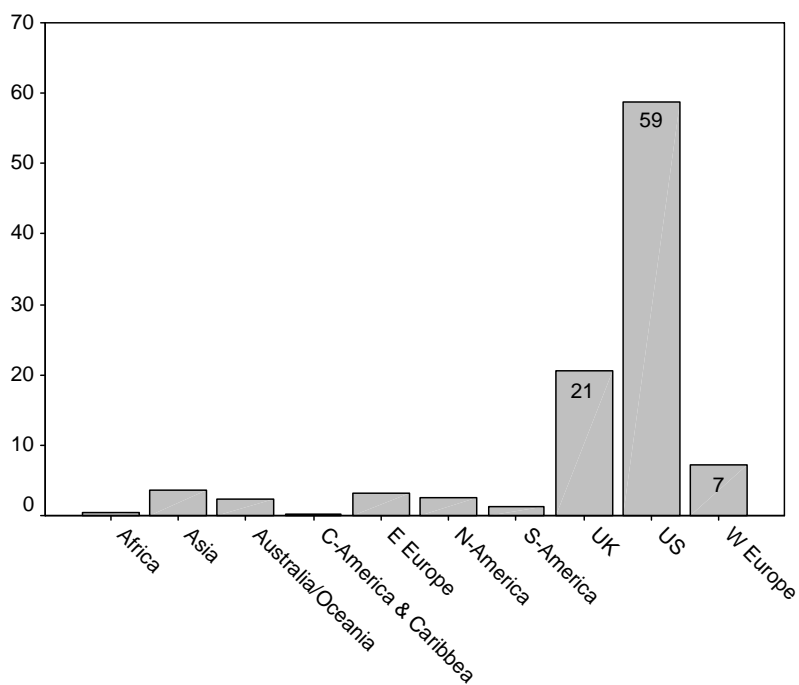
Figure 11 gives the organisation type of the user as given by the DNS registration details. Under a quarter (22%) of sessions were attributed to academic institutions. Most were attributed to either commercial (42%) or net provider (35%) organisations.

Figure 11: Organisation type of user as given by DNS registration details
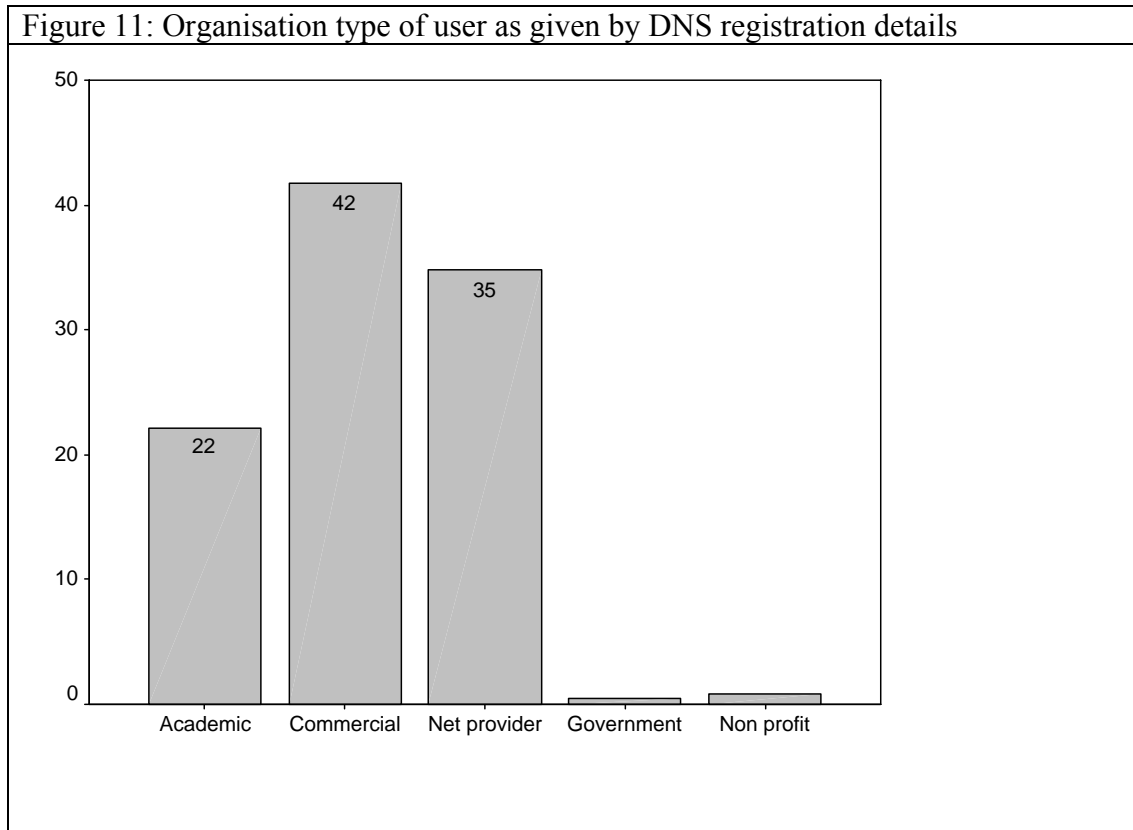
Figure 12 gives the list, top 30, of user academic organisation codes. The most important by some margin is Oxford (ox). About a quarter (24.9%) of sessions are attributed as coming from that source.

Figure 12 The top 30 user academic organisation DNS codes

Academic code	Number of sessions	Percentage of academic sessions
ox	27706	24.9
cam	2436	2.2
nottingham	1887	1.7
uea	1883	1.7
unimelb	1729	1.6
ucl	1702	1.5
dundee	1289	1.2
leeds	1202	1.1
le	1173	1.1
bham	1131	1.0
soton	1011	.9
mmu	996	.9
gla	957	.9
bris	955	.9
bton	905	.8
uu	896	.8
york	847	.8
kcl	845	.8
man	745	.7
hw	697	.6
shef	686	.6
ex	666	.6
ed	615	.6
dur	598	.5
uni-leipzig	573	.5
virginia	572	.5
open	571	.5
glam	558	.5
shu	554	.5
		50.8

In looking at the distribution of type of user over location (US and UK) it can be seen that while over half (61%) of UK user sessions were academic only 6% of US user-sessions were from academic institutions. US commercial users, who made up 45% of sessions here, were btcentralplus (19.4%), AOL (18.8%), btopenworld (4.7%) and many users of these organisations will have been used by UK users to access the Internet. In terms of UK-classified organisations internet access facilities were provided by Cable and Wireless (54%), blueyonder (23.3%), demon (4.9%).

Figure 13: Distribution of type of user over location (US and UK)
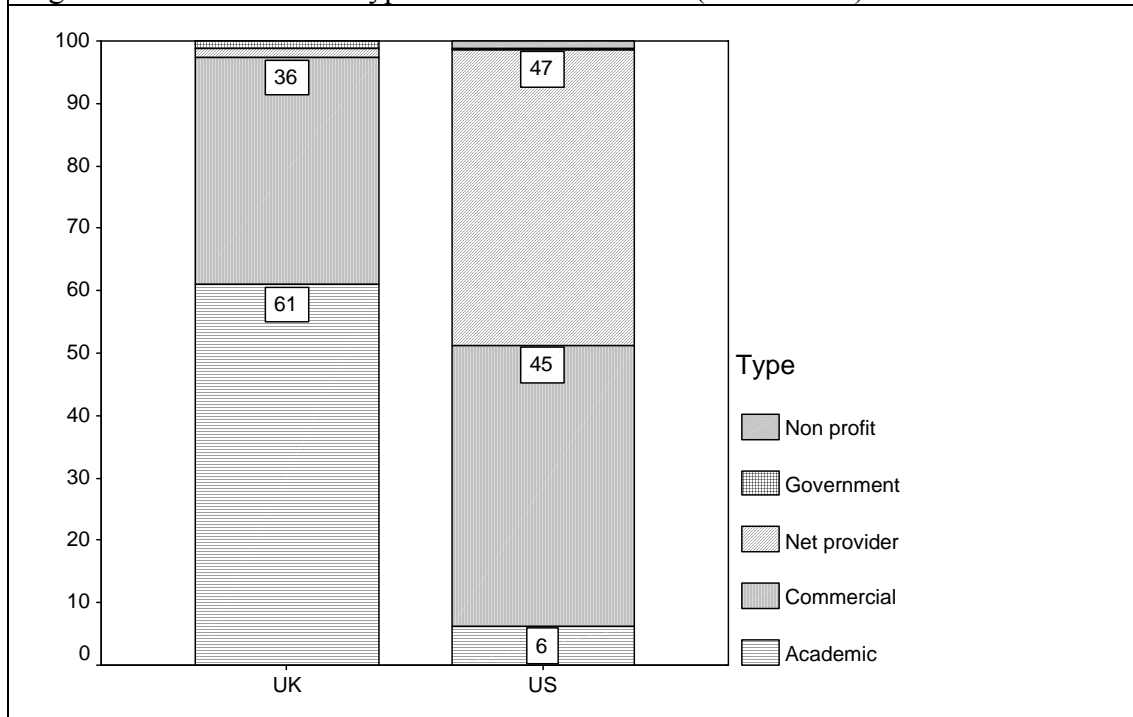
Referrer links were only recorded for 49.9% of sessions. For a third of all sessions (but 69% of sessions where a referrer was identified) the user accessed the Humbul site via a search engine. In terms of search engines used Yahoo made up about two thirds (64.3%) followed by Google (28.5%), Altavista (2.6), BBC (2.1%) and MSN (1.5%).

The following tables the distribution if a search engine was used by type of user (as recorded in their DNS). Where a DNS lookup or a referrer link was not found, the evidence has been excluded. Academics were least likely to use a search engine; but just over half did. Users coming in via a net-provider were most likely to have done so (76%). About two thirds of the 'commercial users' accessed the site via a search engine.

Figure 14: The percentage share distribution if a search engine was used by type of user by DNS registration.
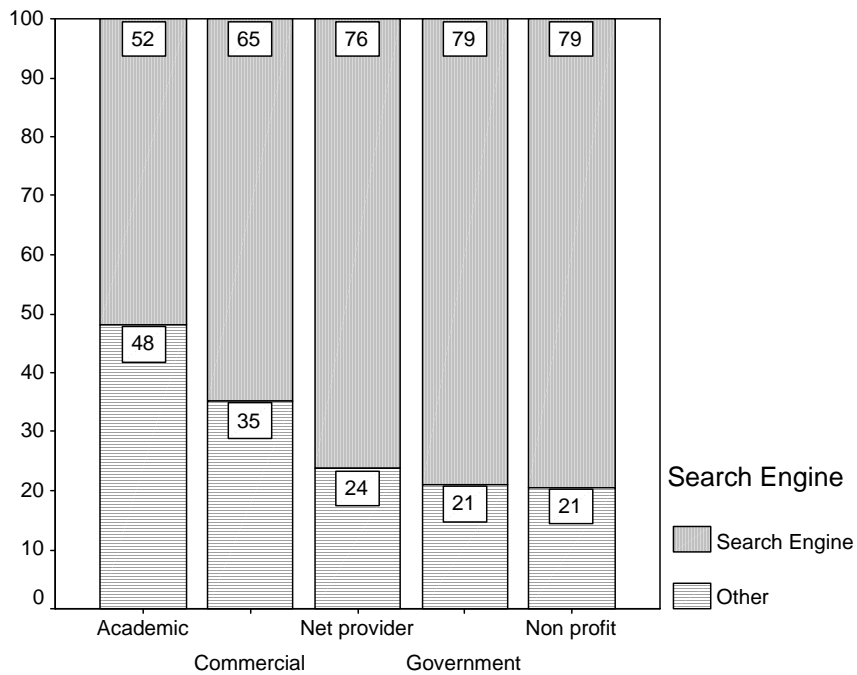
Figure 15 gives the same information but over a regional location. The UK had the lowest use of search engines, with just 45% of UK users accessing the site using a search engine. This is precisely what one would expect, however, since the UK also had the highest usage by academic users.

Figure 15: The percentage share distribution of if a search engine was used by country of user by DNS registration.
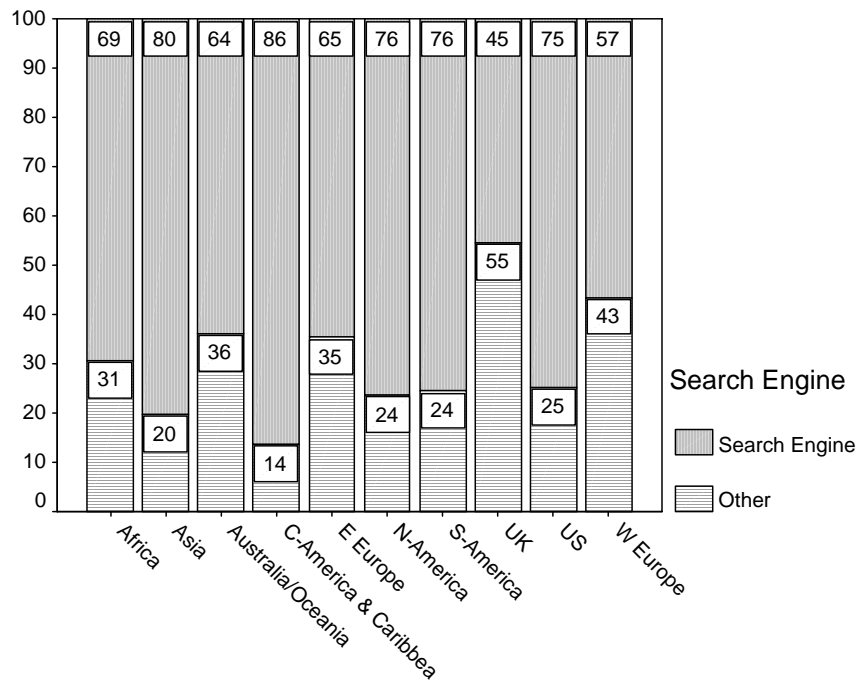
In terms of the number of pages viewed in a session, about 38% of sessions viewed more than one page; an estimated 25% viewed 2 to 3 pages; 10% 4 to 10; and 3% 11 or more pages in a session.

Figure 16 gives the percentage distribution of the number of pages (grouped) viewed in a session.
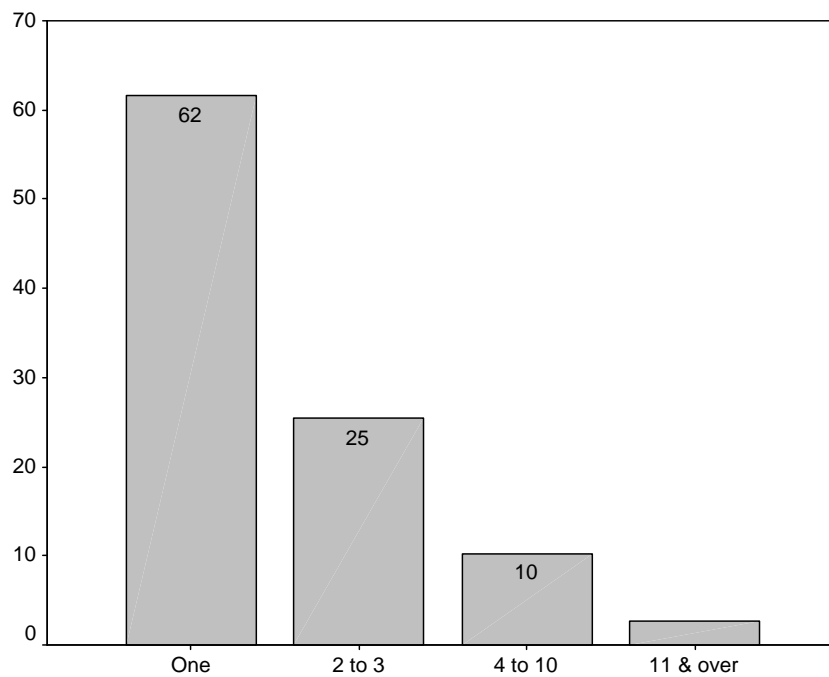
Figure 17 gives the distribution of views in a session by DNS organisation type of user. Net (63%) and commercial users (60%) were most likely to view just one page while academic users (54%) were least likely to do so. The web-logs confirm, in other words, what one would have expected: academic users tended to be the more serious users.

Figure 17 the percentage distribution of views in a session (grouped) by DNS organisation type of user
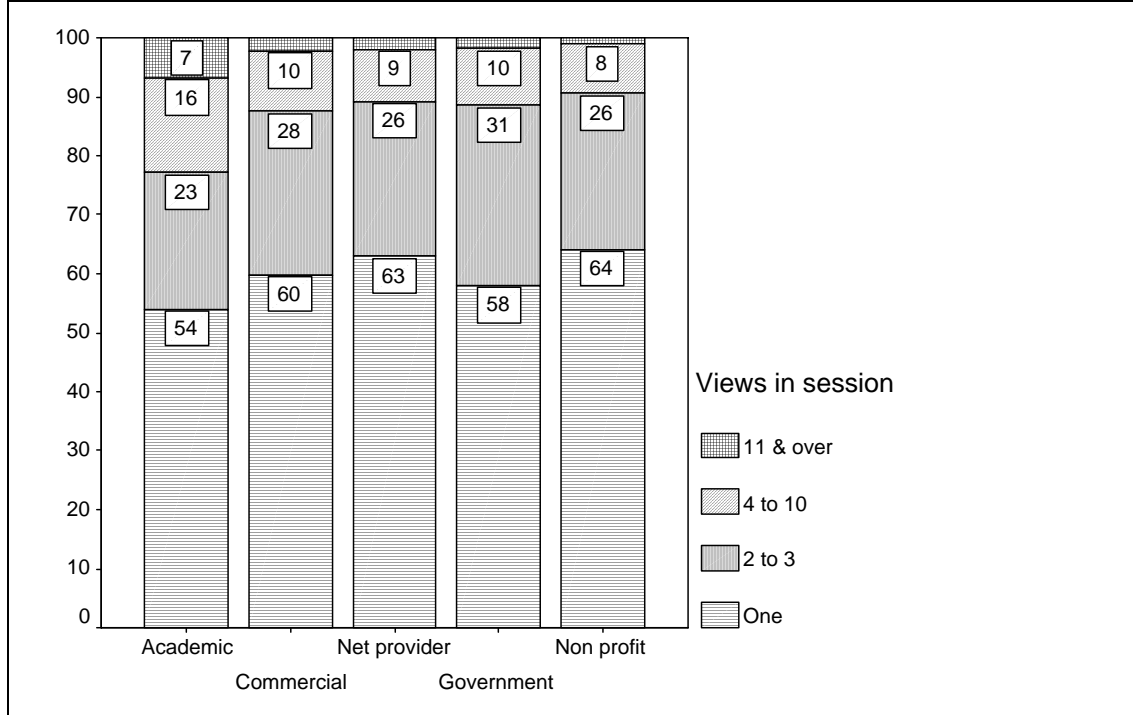
Figure 18 gives the distribution of views in a session, measured in terms of users who had accessed the site via a search engine during the session. Those viewing just one page (73%), or just 2 to 3 pages (74%), were more likely to have come in via a search engine. Sessions where more than 4 pages were less likely to have accessed the site using a search engine. It is possible, of course, that these sessions consisted of viewing more menu pages rather than penetrating to the resources in the collection. However, it is also clear that some users were coming in via a search engine and just browsing one or two pages, and then leaving again.

Figure 18 The distribution of views in a session by if the user had used or accessed the site via a search engine during the session
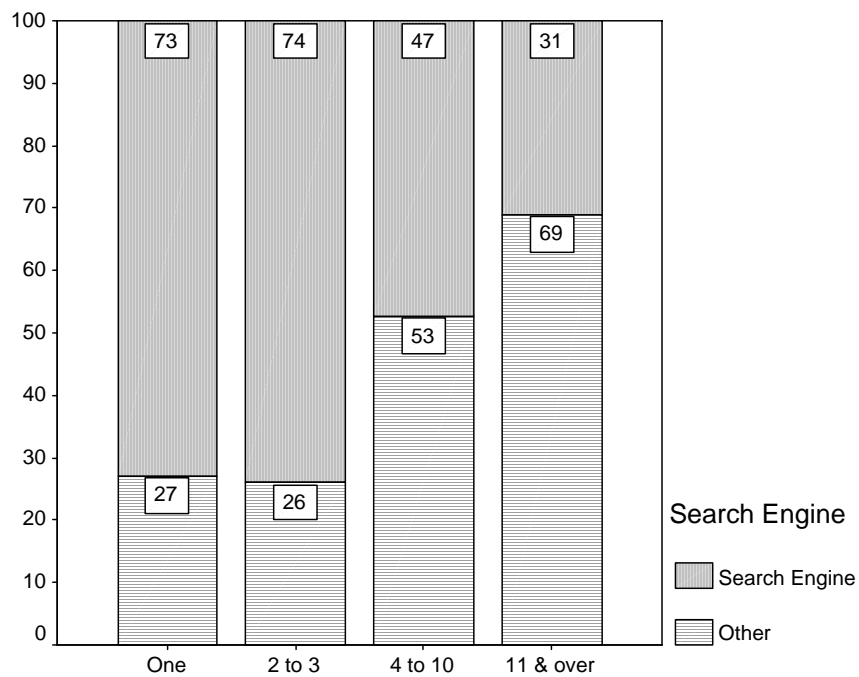
Figure 19 gives the distribution of views in a session by DNS country type of user. US users (63%) are most likely to view just one page; UK users (52%) were least likely to do so.

Figure 19 the percentage distribution of views in a session (grouped) by DNS country type of user
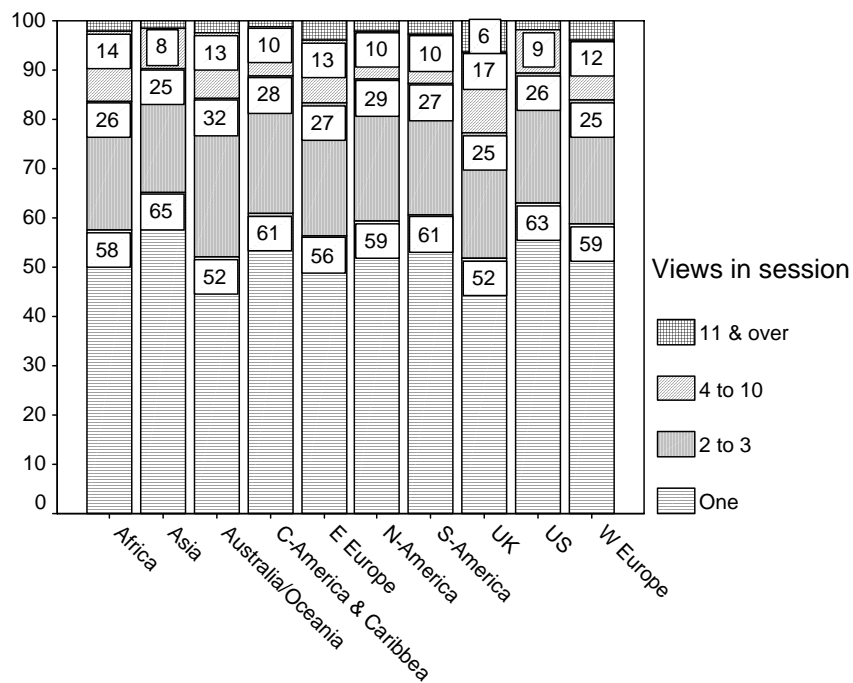
Figure 20 gives the percentage distribution of session time, grouped by DNS organisation type of user. Academics were recorded as having longer sessions. 31% had sessions lasting over 3 minutes. The comparable figure for commercial session users is 21%, with 19% for net-organisation based sessions.

Figure 20 the percentage distribution of session time (grouped) by DNS organisation type of user.
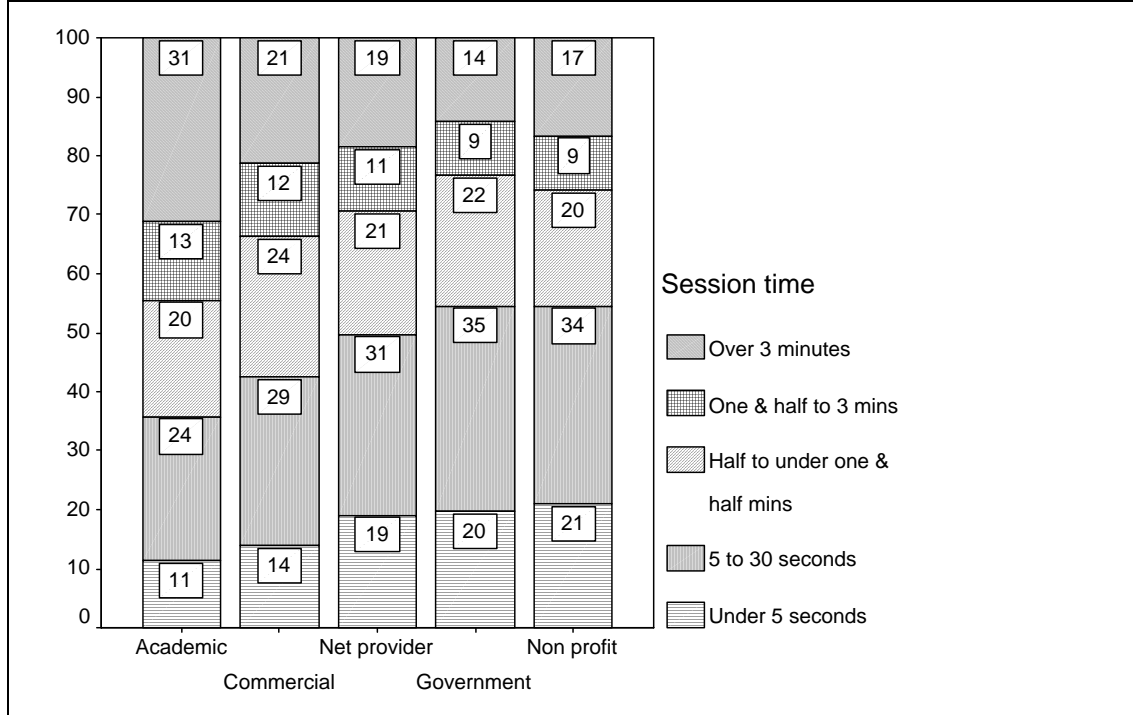
Figure 21 gives the percentage distribution of session time, grouped by whether the user had entered the site from a search engine. Those users not using a search engine were far more likely to have longer sessions. Twenty-three percent had sessions that lasted 3 or more minutes compared to fourteen percent of users who had entered the site via a search engine.

Figure 21 the percentage distribution of session time (grouped) by DNS organisation type of user.
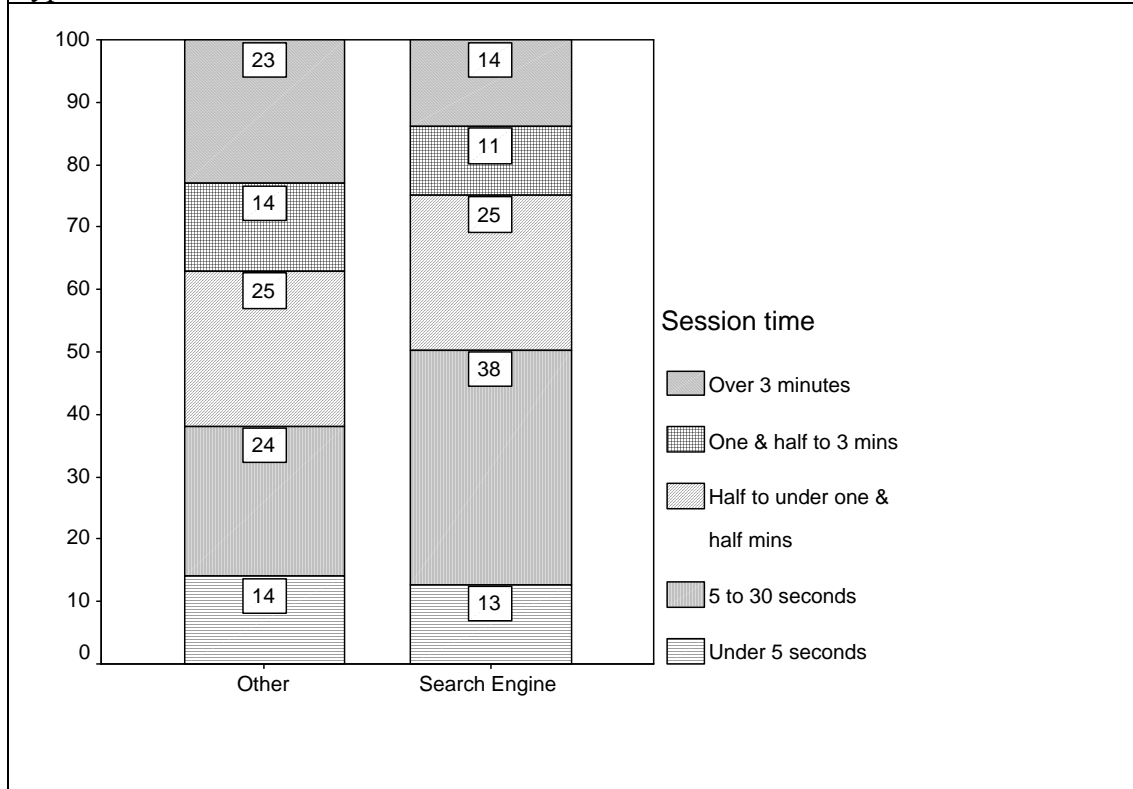
The following (Figure 22) classifies and compares how users navigated their way around Humbul. The key outcome variable here was if the user had accessed an ID (extended summary page), or if the user had clicked on a URI (resource link). The user could find this information using one of three methods: a) a search engine (such as Yahoo or Google); b) the on-site search facility; or c) the site menus - or a combination of these three. A 'menu-user' is defined here as a user who had viewed a sub (menu2) level menu at least once. This grouping of navigation accounted for about half of the sessions. However, about 12% of sessions had just visited the homepage and did not go on to view any subject menu or outcome views. A further 19% of all sessions were 'other' referrer users who also did not view any subject or outcome views.

By examining users' navigational path, we discovered that about two thirds (64%) navigated the site via a search engine; about 12% of sessions used the on-site search facility; 12% used menus; and 11% used some combination of the three method at least once in their session.

Figure 22: Distribution of navigation method
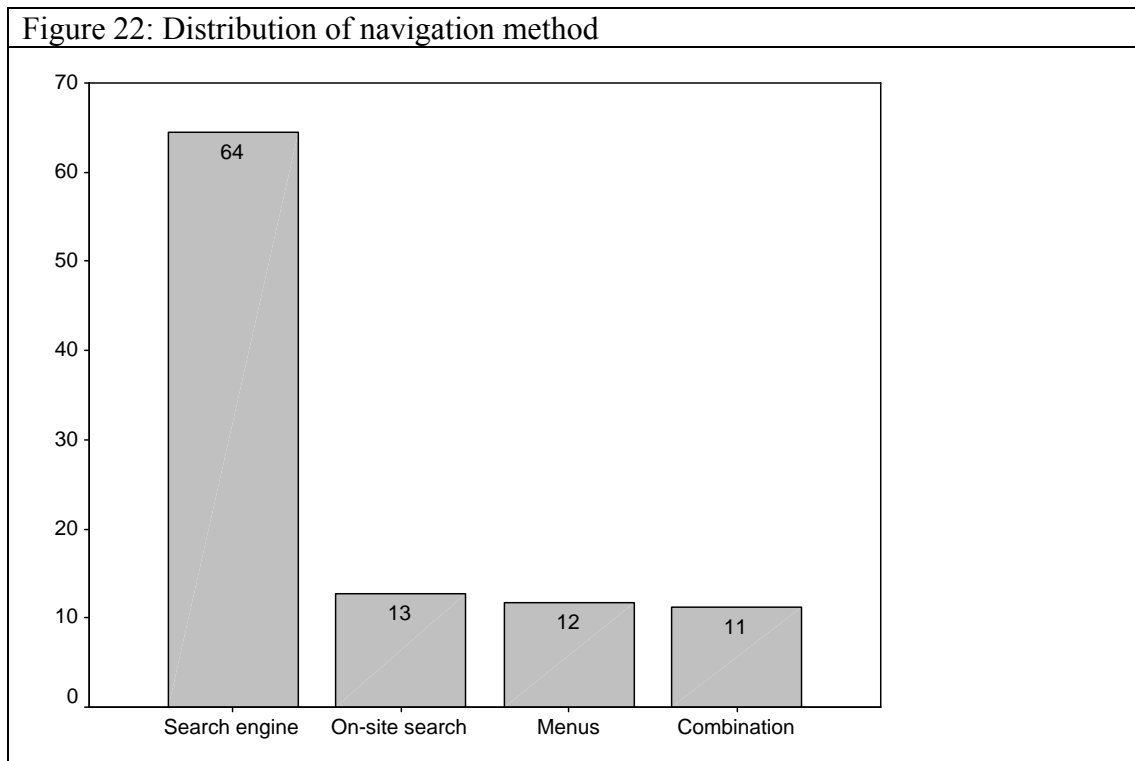
Figure 23 gives the distribution of navigation access by month. There appears to be a greater use of on-site searching and use of a combination of methods between September to December compared to other months.

Figure 23 The distribution of navigation access by month
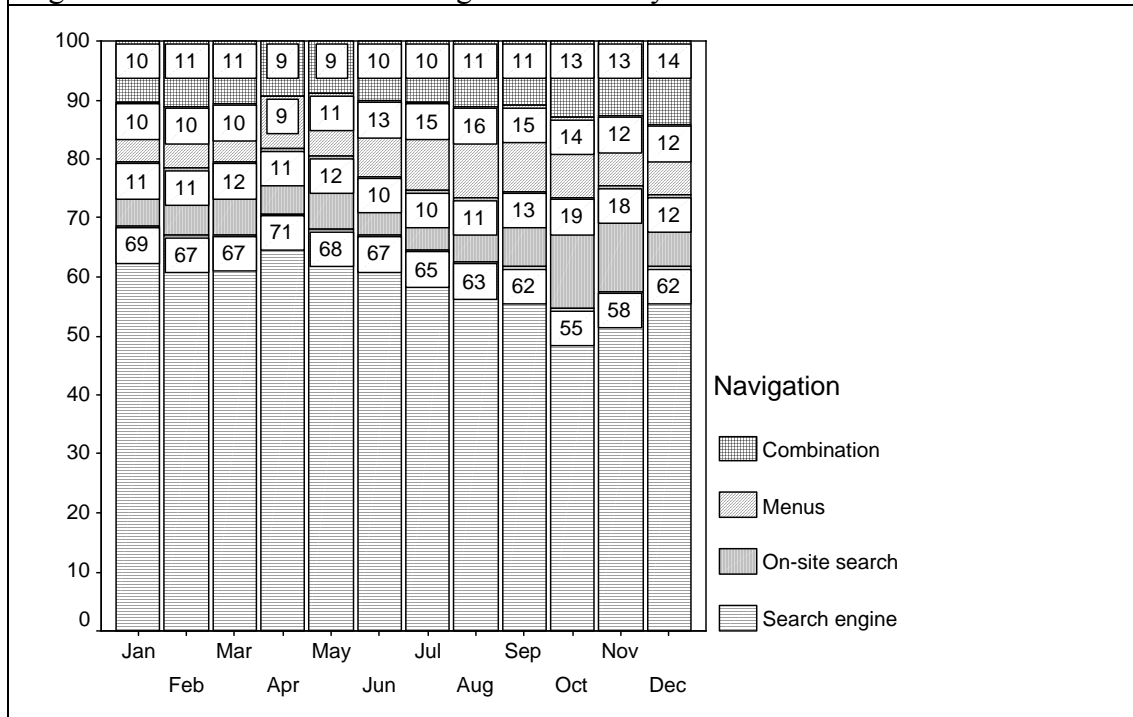
Figure 24 gives the distribution of navigation access by organisation type of user session. Academic users (c.38%) were least likely to navigate using a search engine as compared with 63% of commercial user sessions and just under three quarters (73%) of net type sessions. Academic users were much more likely to use the on-site search facility (29%), menus (18%) or a combination of methods (15%).

Figure 24 Percentage distribution of navigation access by organisation type of user session.
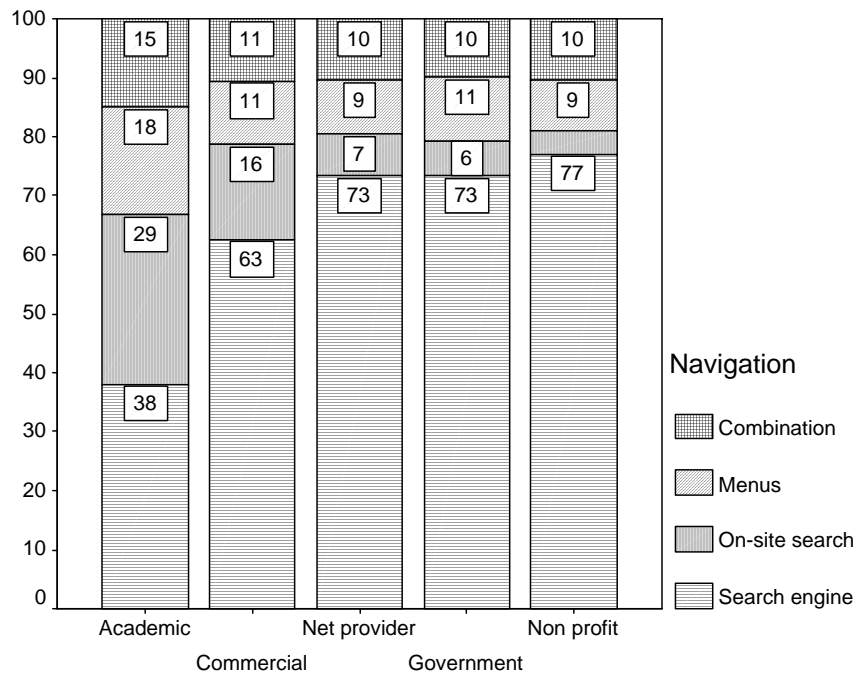
Figure 25 gives the percentage distribution of number of pages viewed in a session across navigation access. Those users coming into the site by a search engine were far more likely just to view one page and leave 66% did so compared to 22% of sessions where the on-site search facility was used or 33% of menu users.

Figure 25 The percentage distribution of number of pages viewed in a session across navigation access.
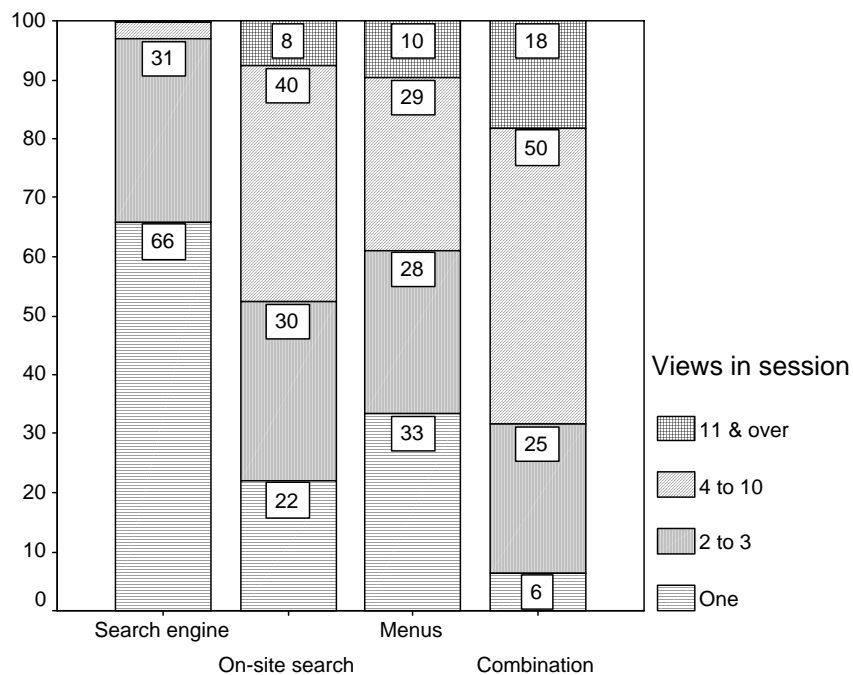
In terms of accessing the item ID extended summary, 84% of those who used a search engine to access the site, also accessed an extended summary. About 20% of those using the on-site search facility accessed an extended summary, 13% of menu-users did so and 48% of those using a combination of access methods accessed such items.

Figure 26 The percentage distribution of ID extended summary items viewed and cross navigation access (sessions)

Search engine	Yes: 84, No: 16
On-site search	Yes: 20, No: 80
Menus	Yes: 13, No: 87
Combination	Yes: 48, No: 52

Extended summary
Yes
No
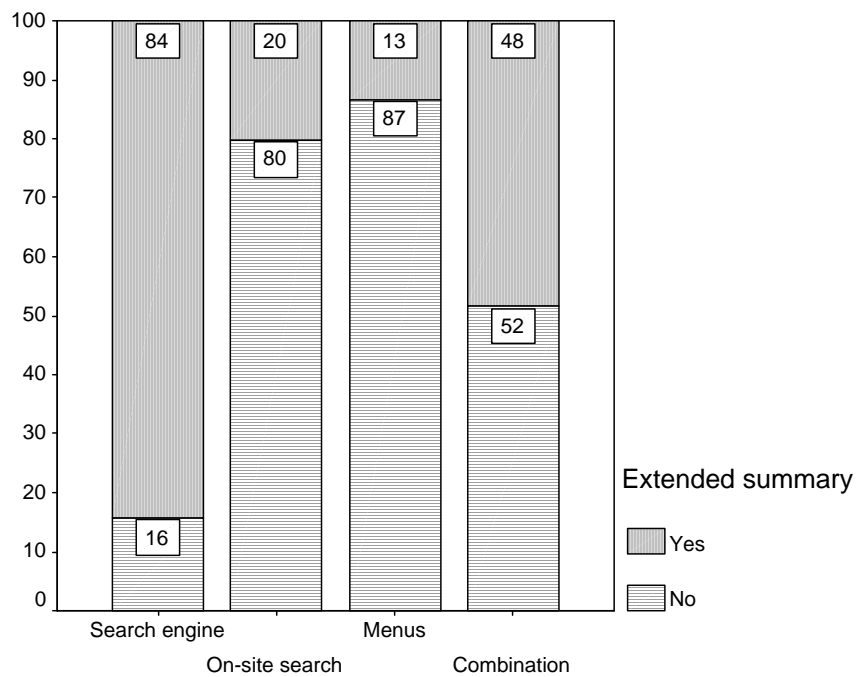
In terms of accessing a URI link, about a quarter (25%) of those using a search engine went on to link to an external resource, a quarter (26%) of those using the on-site search facility did so. About 1 in 5 of users navigating via menus went on to link to a resource and about 39% of those using a combination of methods did so.

Figure 27 The percentage distribution of URI link resources used by navigation access. (sessions)
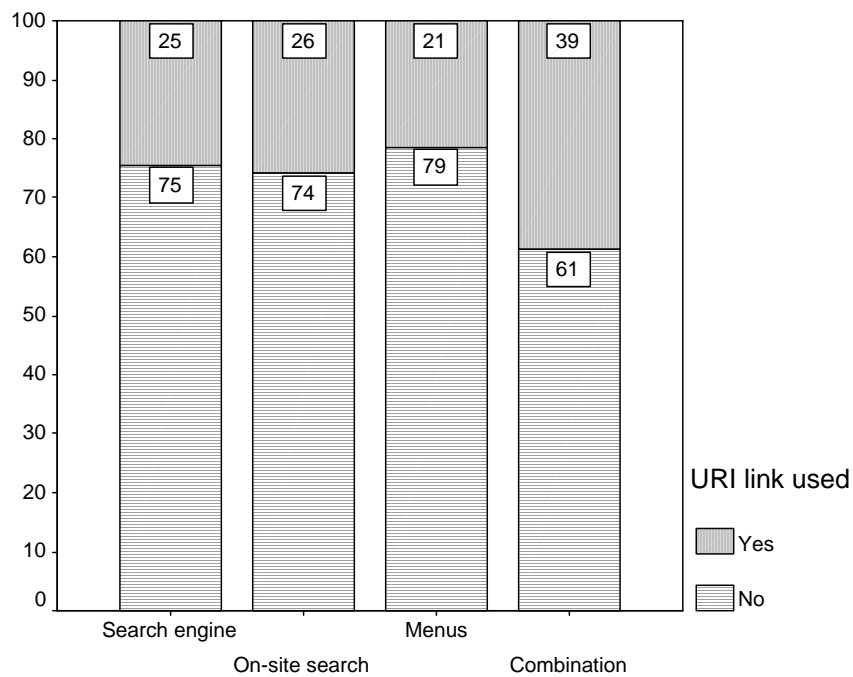
The following gives the distribution of subjects (first subject viewed) viewed (menu users only) in a session. History (25.6%) attracted the most use, followed by English (16.8%), Religion (6.6%), Humanities_a (6.3%) and Philosophy (5.6%).

Figure 28: Distribution of first subject viewed (sessions)
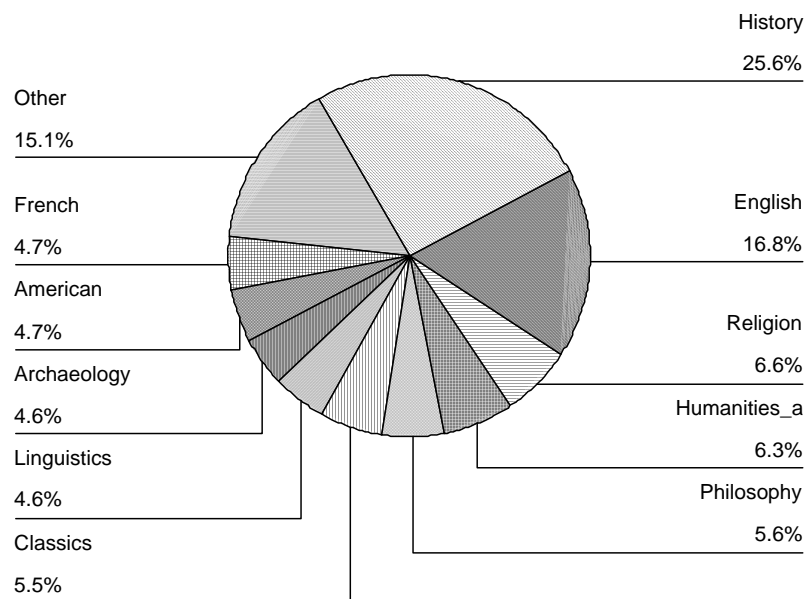
Figure 29 gives the number of items viewed in a session across subject. Those viewing Humanities_a tend to view more items in a session compared to other subjects.

Figure 29 The number of items viewed in a session across subject.

	History	English	Religion	Humanities_a	Philosophy	Classics	Linguistics	Archaeology	American
11 & over	15	15	16	16	17	18	18	22	23
4 to 10	39	39	42	51	42	39	41	41	40
2 to 3	24	25	22	27	22	24	20	20	26
One	22	21	21	6	18	19	22	18	10
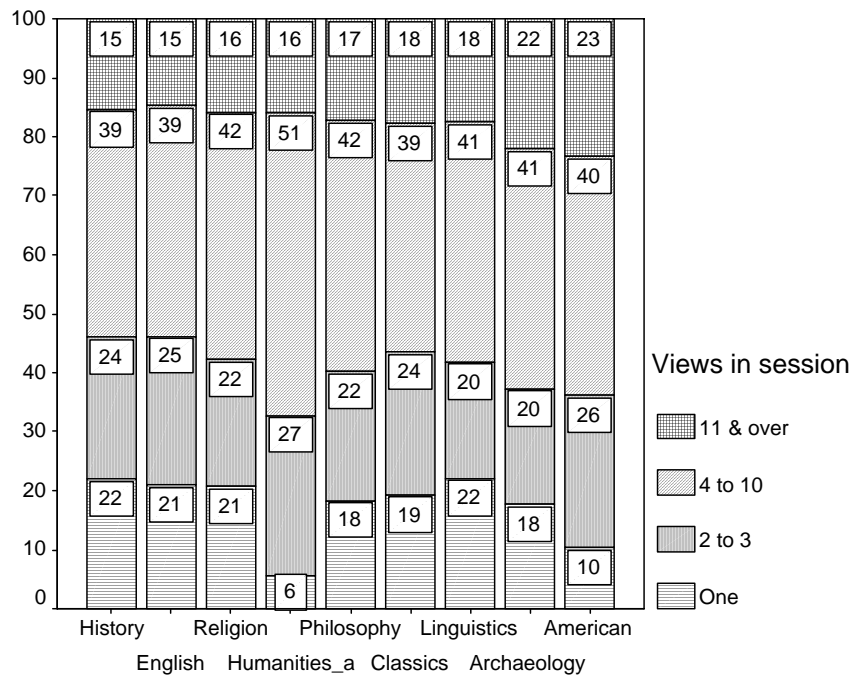
Figure 30 gives the number of extended items viewed across subject. Archaeology (19%) and American (18%) studies attract a greater percentage views to ID extended summary pages and Humanities_a (8%) the least.

Figure 30 The distribution of extended items viewed across subject. .
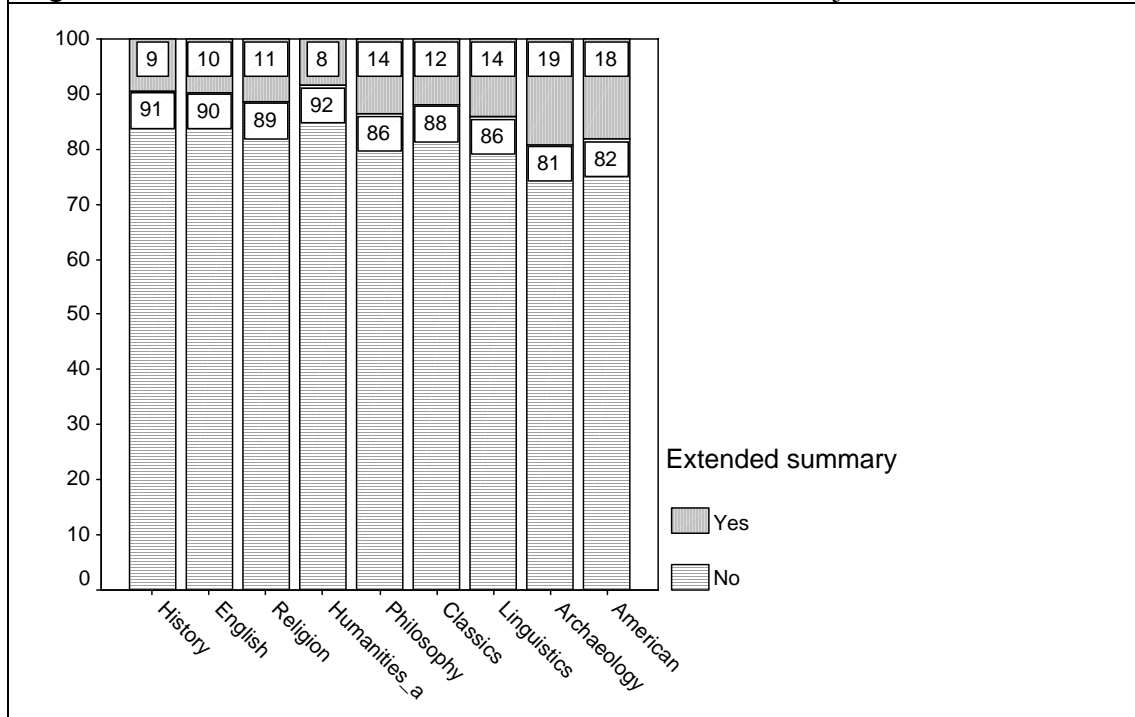
The following table rank lists the first word of Yahoo search expressions used. Common terms such as "the" were excluded. History seems a popular search word to include and 5% of search expressions included history as the first term in a search expression.

Figure 31: Yahoo search words – first search word used only

Term	count	Percentage of all words
History	5803	4.4
world	1037	.8
History	994	.8
English	775	.6
ancient	724	.5
philosophy	622	.5
roman	558	.4
victorian	557	.4
women	544	.4
find	539	.4
john	530	.4
British	525	.4
journal	515	.4
online	505	.4
Russian	461	.3
spark	461	.3
medieval	455	.3
pictures	446	.3
American	427	.3
language	416	.3
		12.8

A5.3.3 Examples of User-Behaviour

It would require a much more extensive analysis than was warranted by the limited user-information at our disposal to undertake an examination of individual user-behaviour. Here, we simply look at the three transactions by way of example.
The first (user-visit 1) is of an IP number that was recorded as visiting the site twice, once on 7 March 2005; then again on 10 October 2005. In March, this user viewed 5 pages. Via current procedures, we were not able to identify this user's domain name server (DNS) details. The user accessed the Humbul site via an organisation called www.netaddress.com and accessed an ID page. One minute and nine seconds later, the user looked at the Slavonic 'sub-menu' page. This is a menu-page listing resources in a reduced summary, providing a link to an extended summary, and then linking to the external resource. Sixteen seconds later, the user completed an internal search using the words 'poet poem poetry'. Nine seconds later, the user returned to the 'sub menu' Slavonic page. We infer that the user was looking for material relating to poetry in Slavonic languages, or relating to Slavonic subjects. We must also infer that the search did not initially provide anything of interest. Twenty-nine seconds later, the user revisited the 'id' page, once more visiting it via www.netaddress.com suggesting that the user had left the Humbul site and then revisited it once more. In all, this user-session lasted about two minutes. The user returned to the site about 7 months later on 10 November 2006. This time, the user just viewed one ID extended summary-item which

they had found on the site via Yahoo and used the search terms 'women's work out' to find the document. The user just viewed one document and left.

Figure 5.32: user-visit (1)				
	(DNS unknown)			
07-Mar-05	14:49:52	id	6365	www.netaddress.com/
07-Mar-05	14:51:01	sub	Slavonic	www.Humbul.ac.uk/
07-Mar-05	14:51:17	search	Poet poem poetry	www.Humbul.ac.uk/
07-Mar-05	14:51:26	sub	Slavonic	www.Humbul.ac.uk/
07-Mar-05	14:51:55	id	6365	www.netaddress.com/
10-Oct-05	21:56:26	id	10759	mx.search.yahoo.com/

The example graphically illustrates the frustrating inconsequentiality of web-log analysis without supporting user information. We might tentatively infer that this user did not find anything significant from their first search; but that the site had achieved some 'recognition' for them to revisit it during a later search. In neither case, can the resource-discovery 'experience' be described as very 'rich'.

The second example (user-visit 2) is of a user who accessed the site on 2 May 2006. The user landed at an ID extended summary page, having found the site using Google. They had used the search terms 'allison pompeian households'. A second afterwards, the server delivered a page labelled '404.html'. A 404 code-page is normally one that informs the client that the page or item had not been found. Traditionally, web-item counting software identifies the 404 coded items in the status field of the logs and deletes these from the count. In our analysis, however, since the 404 item is delivered as an html coded page the analysis will count two items as viewed, even though (in terms of resource discovery), the site visit had yielded nothing by way of information.

Figure 33: user-visit (2)				
02-May-05	17:24:22	id	13518	www.google.com/
02-May-05	17:24:23			

In a final example, a user visited three times, each time using Yahoo to do so. On 17 March 2005, the user found an extended summary document via the search expression 'voices from gaps women writers of color'. The user did not view any other pages. The user returned to the site on 22 March 2006, using the search expression 'nikki giovanni biography timeline'. The user left the site, but then returned to the site 45 seconds later using the same search expression in Yahoo. Five seconds later, the user then clicked on the external resource link and left the Humbul site and visited 'nikki-giovanni.com'. The user then returned about three weeks later and visited the site twice, on both occasions

using Yahoo. The first time, s/he used the search expression 'voices from the gaps' and on the second time, 49 seconds later, using the search terms 'voices from the gaps women writers of a color'. On both occasions the user was delivered an ID extended summary page. This user seemed to prefer to use Yahoo rather than the on-site search facility or menus. Perhaps the idea of getting to grips with a menu structure was more daunting than using Yahoo. At all events, we should probably classify this as a user that had 'found' a resource through the subject-portal on this occasion:

Figure 34
user-visit (3)

17-Mar-05	20:32:36	id	4405	search.yahoo.com/
22-Mar-05	18:00:04	id	9676	search.yahoo.com/
22-Mar-05	18:00:49	id	9676	search.yahoo.com/
22-Mar-05	18:00:54	URI	nikki-giovanni.com	www.Humbul.ac.uk/
15-Apr-05	14:57:45	id	4405	search.yahoo.com/
15-Apr-05	14:58:36	id	4405	search.yahoo.com/

A5.3.4 Humbul Web-Log Analysis: Conclusion

Site Usage.

The Humbul site saw, on average, about 6-8,000 items/pages viewed per weekday in an average of 2,500 sessions. About half the users were from the USA. Under a quarter of the sessions could be directly attributed to academic institutions. We may therefore presume that a minimum of c.550 sessions per day were from academics, with a further cohort of UK academics accessing the site via commercial servers – perhaps doubling that number of sessions. Of course, many of these academic visitors may have been for teaching purposes, or undergraduate visitors in search of materials for projects and dissertations. We should not disaggregate teaching and research too clinically in our resource-discovery analysis. Since site revisits appear to be at a low level (though CIBER was able to furnish us with no statistic upon revisits), we may cautiously infer that only a small proportion of the research cohort in the UK, identified as 50-60,000 in **A2** (above), used the Humbul service in 2005 – perhaps in the region of 1-10% with the likelihood that it is in the lower quartile of that range.

Discipline Distribution.

The dominance of History users of the site is even more pronounced than the statistics suggest. If we compare the subject distribution [figure 6] with the subject distribution by RAE2001 returns [**A2**], this is as striking as is the under-representation of Modern Languages and Linguistics, and (to a lesser extent) Philosophy, Law and Religious Studies (Law not being a subject represented in the RDN resource discovery networks).

Site Penetration

Academic users tended to make more serious use of the site when they visited it. They were least likely to 'bounce' out of the site having visited it. An encouraging statistic from the analysis is that 31% of the academic visitors spent over 3 minutes when they visited the site. Only a small proportion, however, used the on-site menus and search engines. The numbers of academic users who accessed an extended summery of a resource is also encouraging. But one of the most resonant conclusions of the analysis is that only a minority of these users went on then to link to an external resource.

A5.4 AHDS Web-Log Analysis

A5.4.1. Overall Site Usage

An overall view of site-usage is provided by the number of 'hits' or views per day. AHDS had approximately 1000 to 3000 views per day. Use is punctuated by occasional high volume day usage 'spikes' that can reach as high as 10,000 views.

Figure 35: Daily usage AHDS February to September 2005
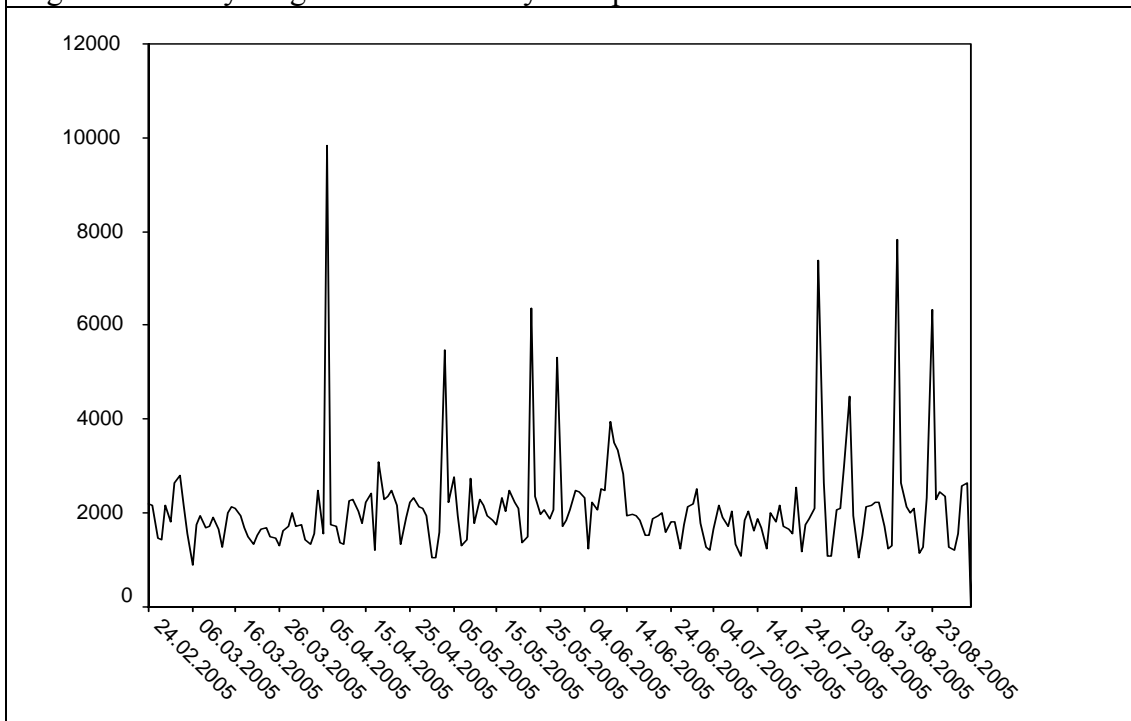
Interestingly, a large percentage of 'hits' seem to have occurred in August – in contrast to the Humbul evidence above. We suspect that this intensity of usage is the result of the long vacation research traffic demands, coupled with MA and MPhil dissertation work. If this is the case we are struck by the fact that the comparable evidence fro Humbul is a trough at the equivalent period (Figure 8).

The way the AHDS site is constructed means many file-names and directories share the same name, irrespective of subject. If a record of subject-usage was to be found then this could only be done at the directory level. But the problem with using a directory-name approach is that there are a number of pages associated with a subject-directory. So, for example, 'History' has a number of menu-type pages and so a purely directory approach gives a biased view of activity, but perhaps provides an overview of subject popularity.
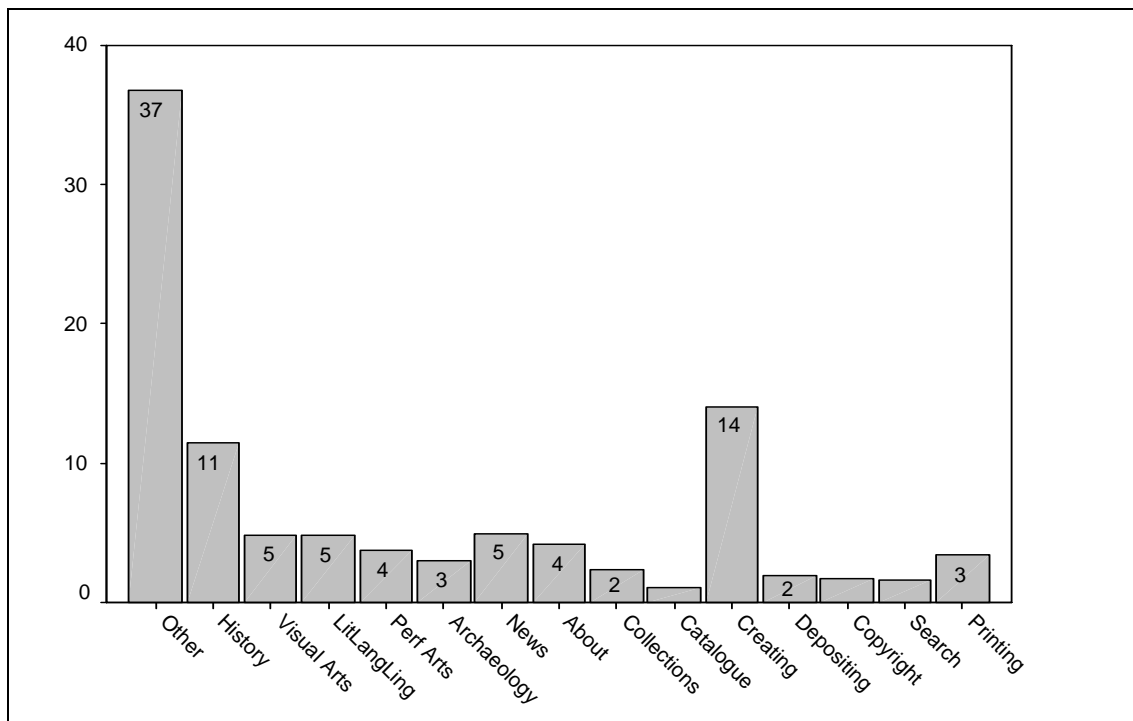
Figure 36: Top level directory usage over March to August for the five subjects

A5.4.2 User Session Analysis

The number of user sessions was 151,998 over February to August - 600 to 900 user sessions a day. This provides the numbers (not 'hits') using the site.

Figure 37: Daily number of user sessions
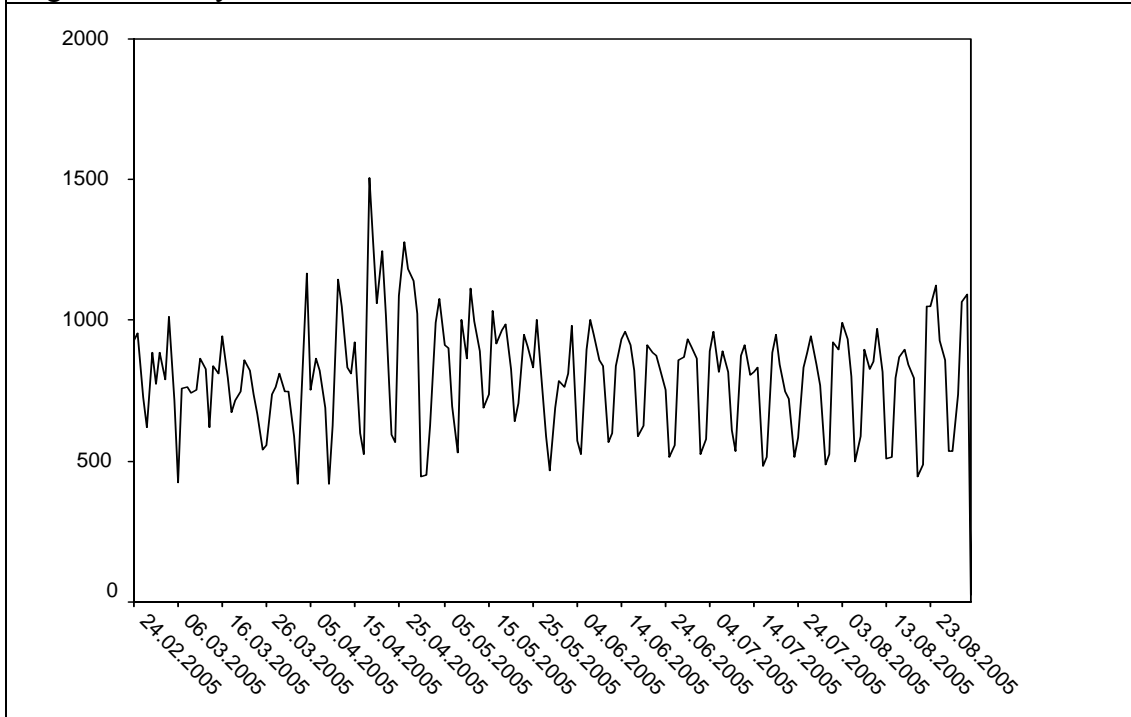
The DNS Analysis of the AHDS site is as follows, with access from 'commercial' DNS dwarfing all other organizational entities:

Figure 38: Distribution of sessions by DNS organisational usage
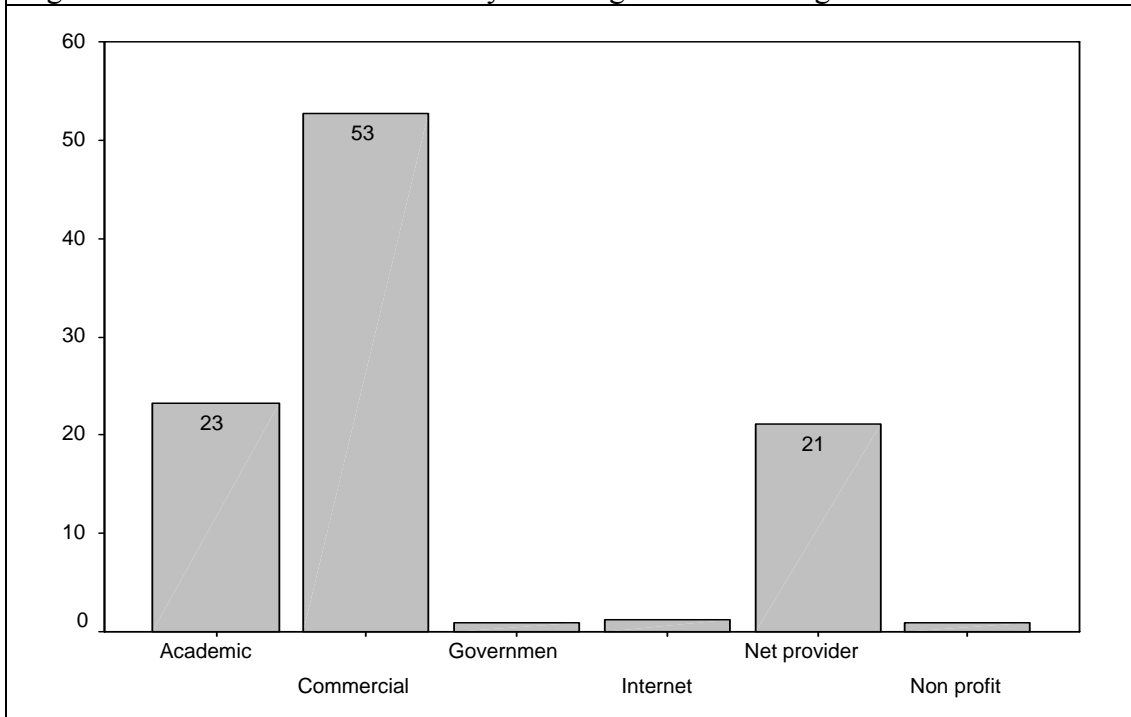
In the equivalent referrer-analysis to that conducted on the Humbul evidence, 21% of the users came to the AHDS via a search engine. There is, however, a high 'unknown user' element to this evidence and whether this should be accounted also as users who also arrived via a search-engine is unclear. It is possible that the true percentage of search-engine derived visitors is 40% or more.

Figure 39: Distribution of sessions over referrer link.
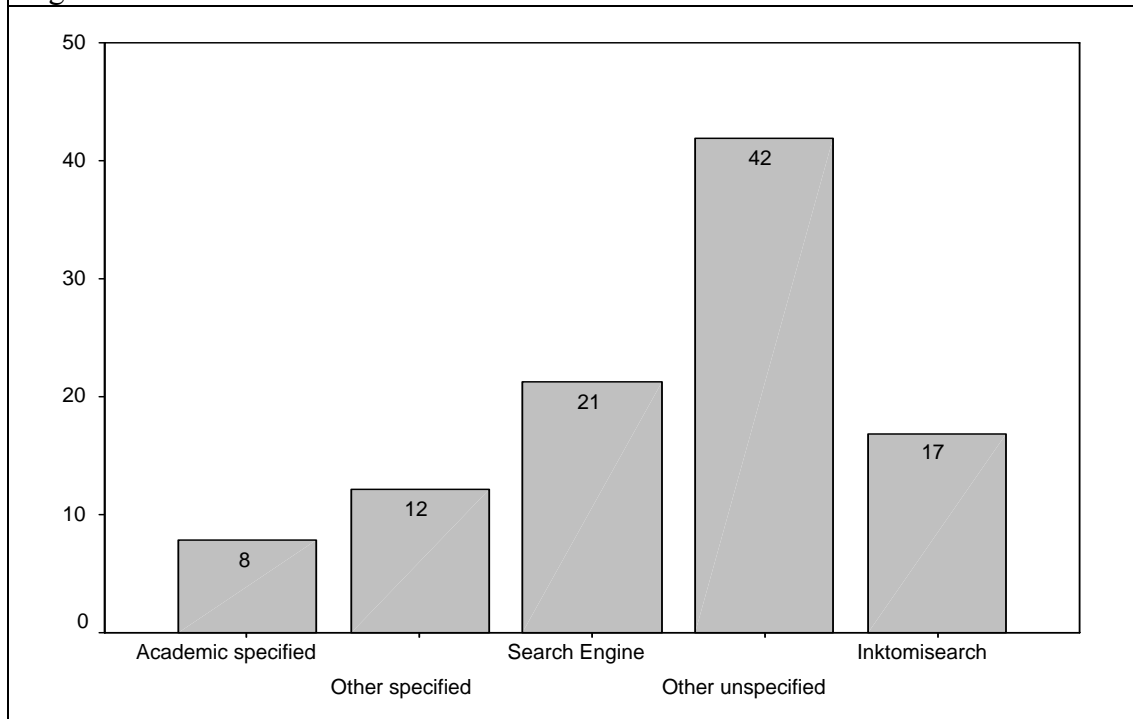
The following table lists the top 15 referrer sites in the group 'other unspecified'. The top 15 accounted for 51% of other unspecified sessions.

Figure 40: Top 15 referrers in "Other unspecified"
www.stumbleupon.com
www.onebird.com
www.ivritype.com
dublincore.org
www.uky.edu
aolsearch.aol.co.uk
www.cmswatch.com
www.tei-c.org
www.library.cornell.edu
www.ifla.org
www.minervaeurope.org
www.dlib.org
www.drh.org.uk
aolsearch.aol.com
www.nla.gov.au

The following table lists the top 15 referrer links in the group 'Academic specified'. The top 15 account for 60% of sessions. We have singled out in green the referrals that may have been significantly affected by internal traffic within the AHDS site. We have also singled out in red the referrals from Humbul and Artifact.

Figure 41: Top 15 referrers in "academic specified"
hds.essex.ac.uk
www.hw.ac.uk
www.data-archive.ac.uk
www.kcl.ac.uk
appserver.pads.arts.gla.ac.uk
census.data-archive.ac.uk
edina.ac.uk
www.Humbul.ac.uk
www.Artifact.ac.uk
www.lib.cam.ac.uk
www.jisc.ac.uk
www.esds.ac.uk
www.bodley.ox.ac.uk
www.ukoln.ac.uk
www.abdn.ac.uk

The next figure gives the distribution of referrer link by DNS organisation of user. In terms of academics, 15% entered the site via academic-specified links, 17% via a search engine and 62% via other unspecified.

Figure 42: The percentage distribution of referrer link by DNS organisation of user
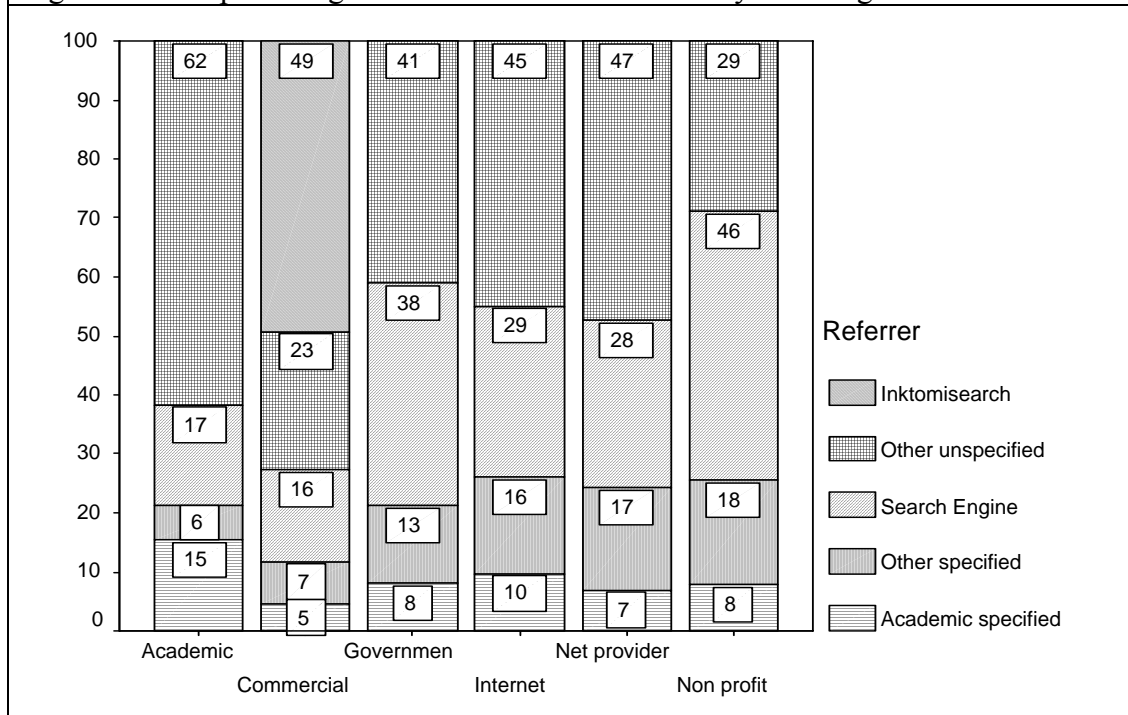
The following lists the DNS name of the referrer group 'Other unspecified' (i.e. the area marked 62% in the Academic bar in the above chart). The top 15 organisations accounted for 36% of the total were as follows. Once more, those likely to have been influenced by internal AHDS traffic are highlighted in green:

Figure 43: Top 15 academic institutions identified as referrer ('other unspecified')
pc094-016.odds.kcl.ac.uk
pc094-015.odds.kcl.ac.uk
pc094-026.odds.kcl.ac.uk
linux01.lib.cam.ac.uk
pc094-017.odds.kcl.ac.uk
pc094-010.odds.kcl.ac.uk
morse.ucs.ed.ac.uk
bottle.gla.ac.uk
pc094-030.odds.kcl.ac.uk
dozer.infodiv.unimelb.edu.au
farnham.surrart.ac.uk
pc168-21.UB.UU.SE
atticus.ahds.ac.uk
xena.lib.unimelb.edu.au
dahds7.essex.ac.uk

These 'unspecified' users and specified 'commercial users' are most likely to enter the site using a search engine:

Figure 44: Commercial referrer group (other unspecified & unknown)

rr
5.7%

pol
5.6%

btopenworld
3.2%

btcentralplus
10.2%

blueyonder
5.5%

as9105
4.1%

aol
15.3%

yahoo
2.1%

Other
48.3%
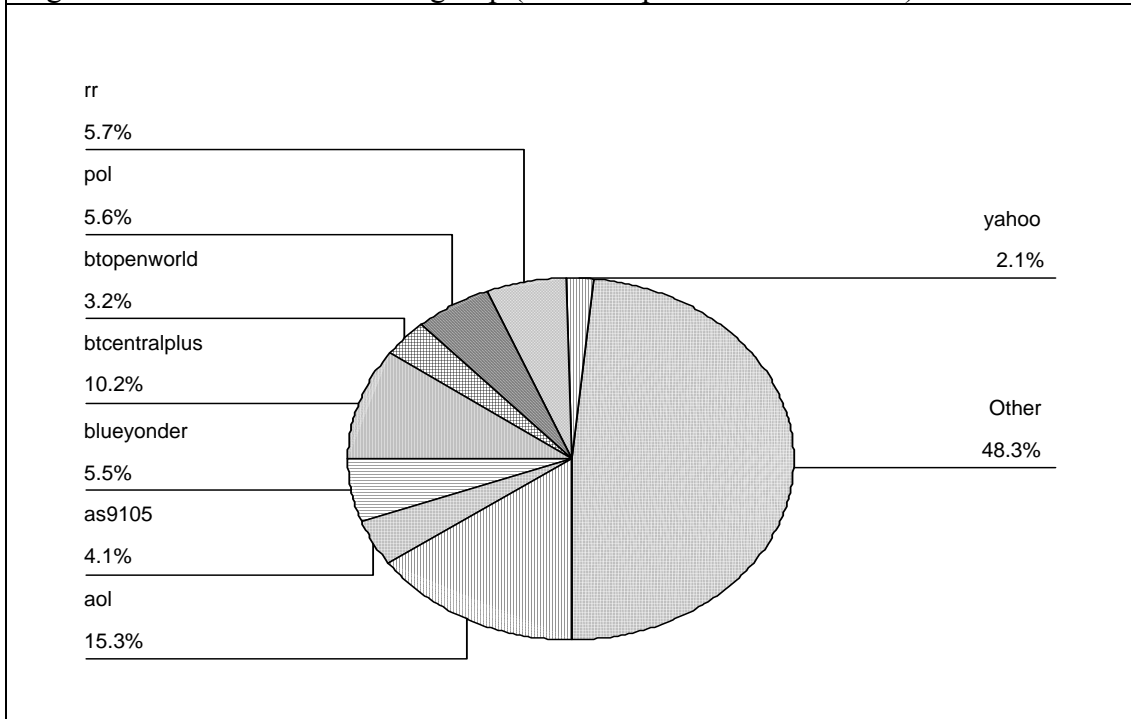
The following gives the country location of academic user sessions only. Academic institutes are said to be less likely to mis-register their location. As can be seen most academic sessions (86%) comes from the UK.

Figure 45: DNS country distribution of user sessions – academic institutions only

Germany
1.2%

Australia
3.8%

Other
2.2%

United States
6.9%

United Kingdom
85.9%

Other: Countries accounting for less than 1% individually
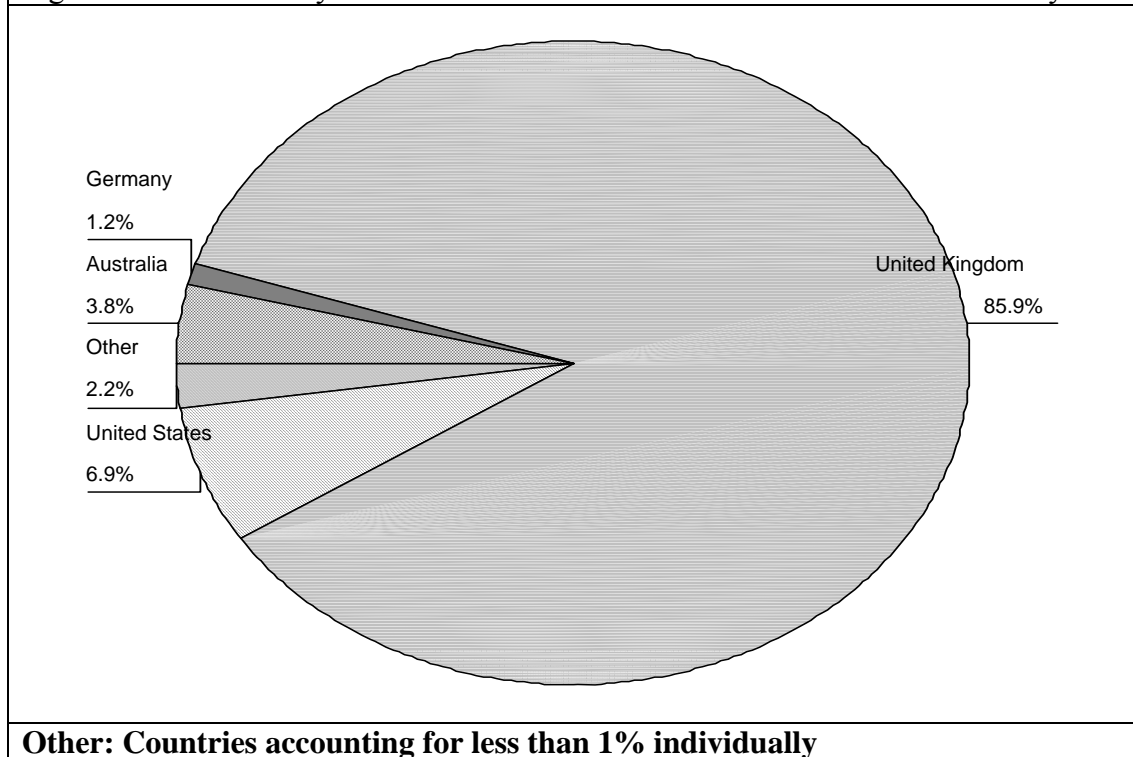
When referrer information is broken down into the 7 highest user institutions it shows that Cambridge (23%), Essex (24%) and Oxford (22%) have a relatively high percentage of users coming in via their university servers. Once again, however, the statistics for Essex may be influenced by internal AHDS traffic.

A5.4.3 Site Penetration

How many pages on the AHDS site were viewed in a 'session'? The percentage distribution of number of pages viewed in a session show few users viewed more than 2 pages in a session. Three quarters (72%) viewed one page. Viewing one page and then exiting is described by CIBER as 'bouncing'. It delineates the user-scenario in which a search engine facilitates information-gathering for a user, who views a site, realises that it is not for them, and leaves. This behaviour is more apparent with search engine/directory listings, where there is little cost in cycling through the first 10 - 20 hits. In terms of the overall referrer-group 78% of search-engine users left after viewing one page. Commercial users were most likely to view one page in a session (71%) whilst academic users were least likely to do so (58%). Those users who viewed more pages were more likely to conduct an internal site-search. Half (56%) of those sessions viewing 11 or more pages did so. 28% of all sessions viewed more than one page, 19% viewed 2-3 pages; 6% 4-10; and 1%, 11 or more.

A5.4.4 Subject Analysis

The AHDS five subject-areas were used as the unit of analysis here: History, Visual Arts, Performing Arts, Literature, Language and Linguistics, and Archaeology. Most user sessions (75%) did not view a subject page. Of those that did, most (23%) just viewed one subject.

In terms of subject, History was the most popular: 38% viewed one page from History; 19% viewed a Visual Arts page; 18% viewed a Literature Language and Linguistics page; 14% Performing Arts; and 12% Archaeology. But the existence of the independent website access for each of the services renders this analysis very tentative – many users of (for example) the Archaeology site will have accessed it directly, and not through the AHDS central server.

Figure 46: Frequency distribution over first subject viewed

Subject	Frequency
History	38
Visual Arts	19
LitLangLing	17
Perf Arts	14
Archaeology	12
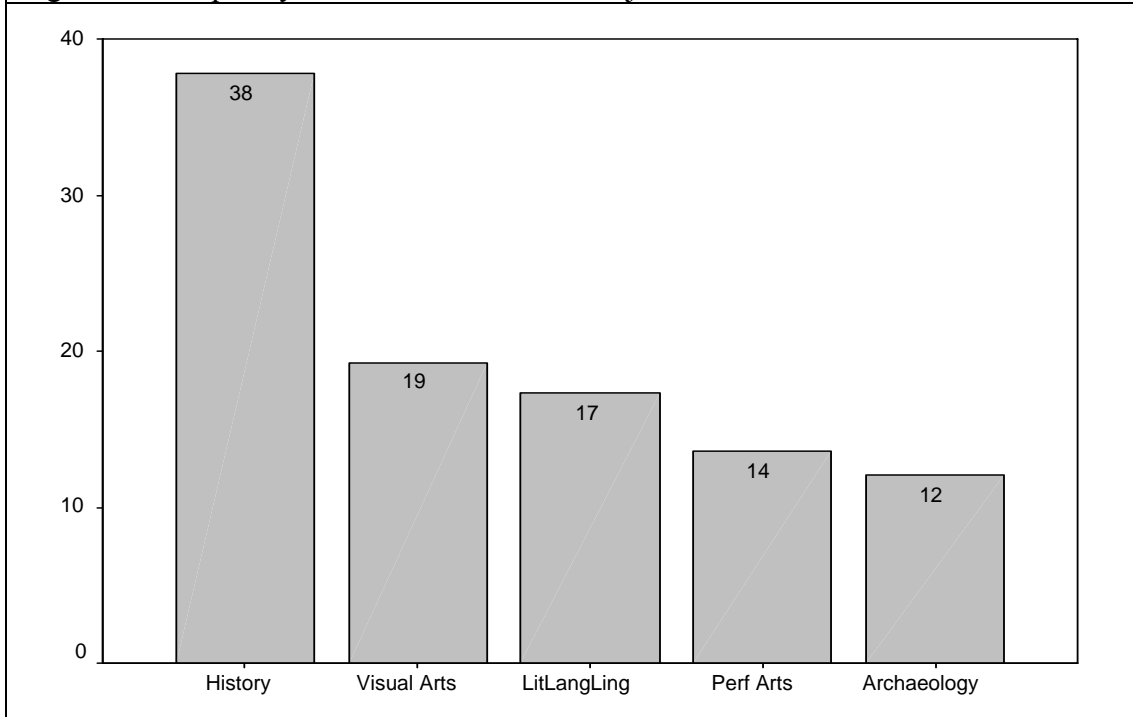
There is a greater use of search-engines by users viewing Literature, Language and Linguistics (33%), as compared with users viewing Visual Arts (15%) and Archaeology (15%).

Figure 47: Percentage distribution share of referrer link by first subject viewed.
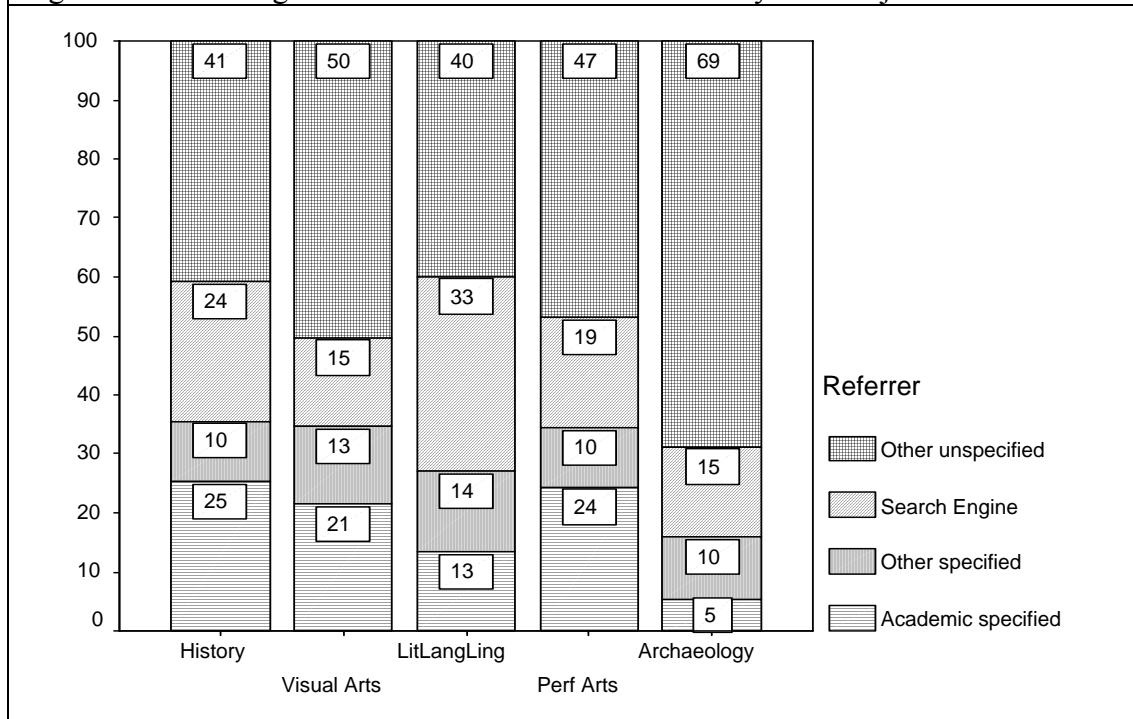
When in a subject-grouping, users could either search the database view, a listing or visit a page listing recent items. The following gives the percentage share of these activities, broken down by subject. It appears that few users go on to search the database. Less than 5% do that in each subject, save for Archaeology, where approximately 33% of users go on to do a search. Of course, it is impossible to establish by this kind of analysis the extent to which site-design is a factor in the behaviour of users at this point.

Figure 48: Percentage share of type of view by subject-grouping .
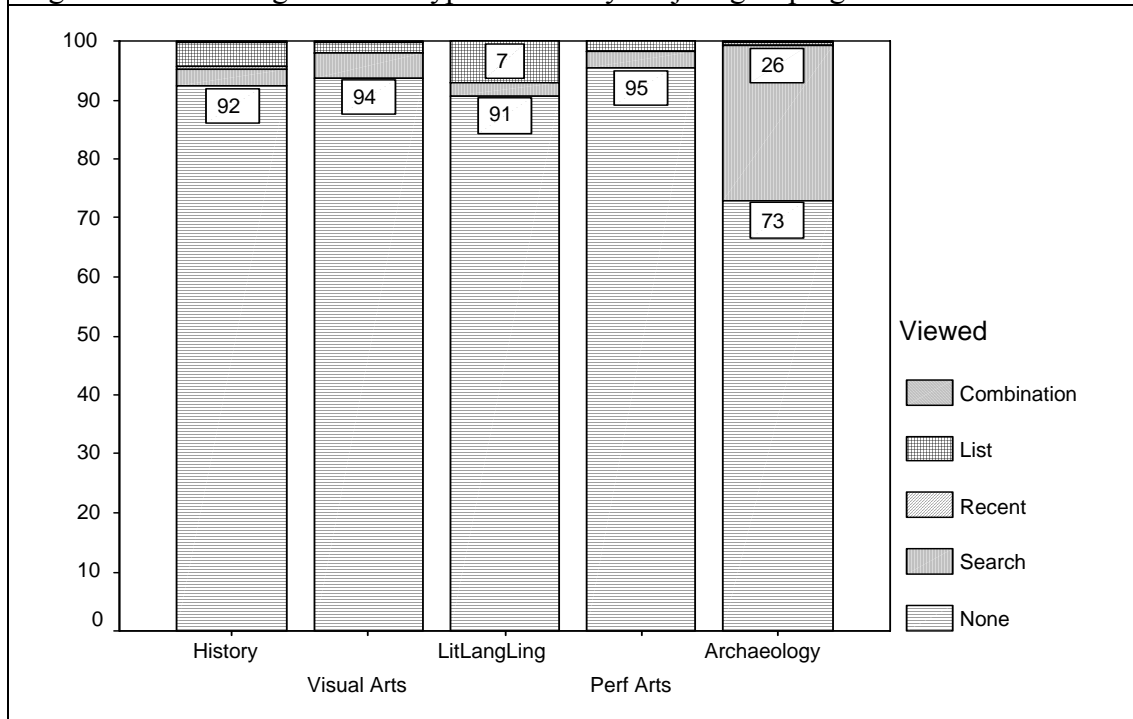
The following gives an idea of the other non-search pages users were looking at. Without a more detailed analysis of the pages in question (difficult to provide because of the diversity of the AHDS site), this is of limited utility since the most frequent category is 'other'.

Figure 49: History – frequency of pages viewed.

Other
46.1%

about
2.3%

census-statistics
5.1%

chcc
4.6%

chccaccess
2.6%

historical-maps
11.1%

nineteencentury-cens
4.6%

staff
10.6%

special-collections
2.3%

projects
10.7%
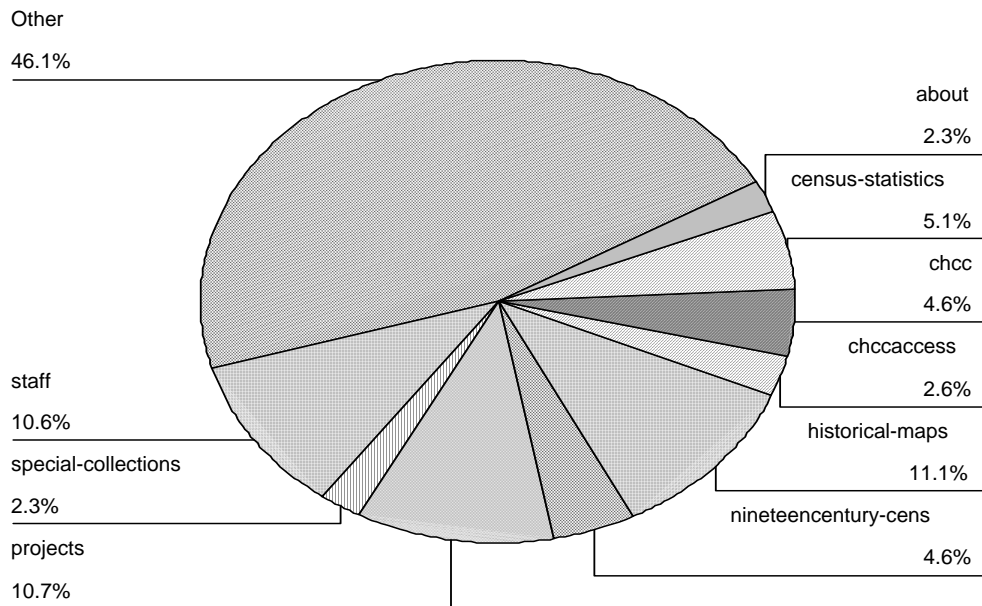
Figure 50: Visual arts – frequency of pages viewed.

Other
88.4%

vads_ag
2.7%

licence
3.2%

guides
2.1%

digitalpicture
3.6%
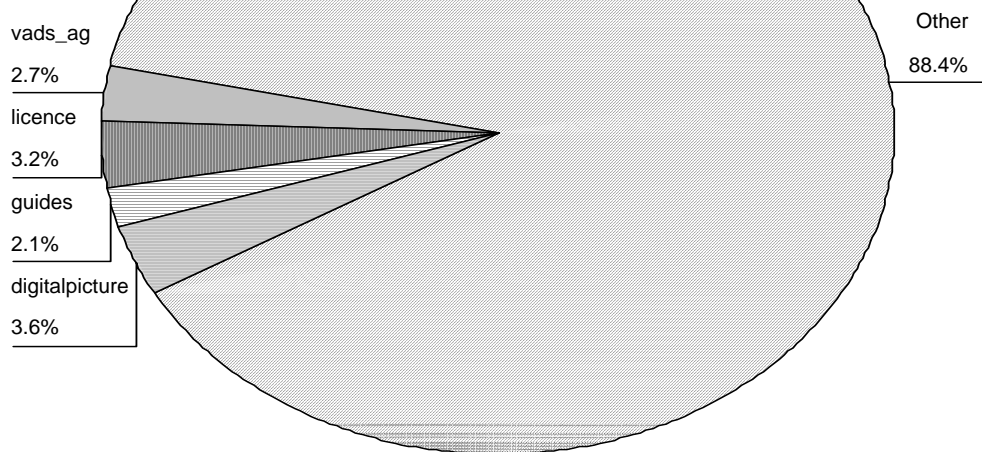
Figure 51: Literature, Language Linguistics – frequency of pages viewed.

changes
3.3%

about
4.4%

ebooks
19.1%

Other
35.7%

FreeEbooks
26.8%

workshop
3.6%

new_deposits
4.1%

staff
3.0%
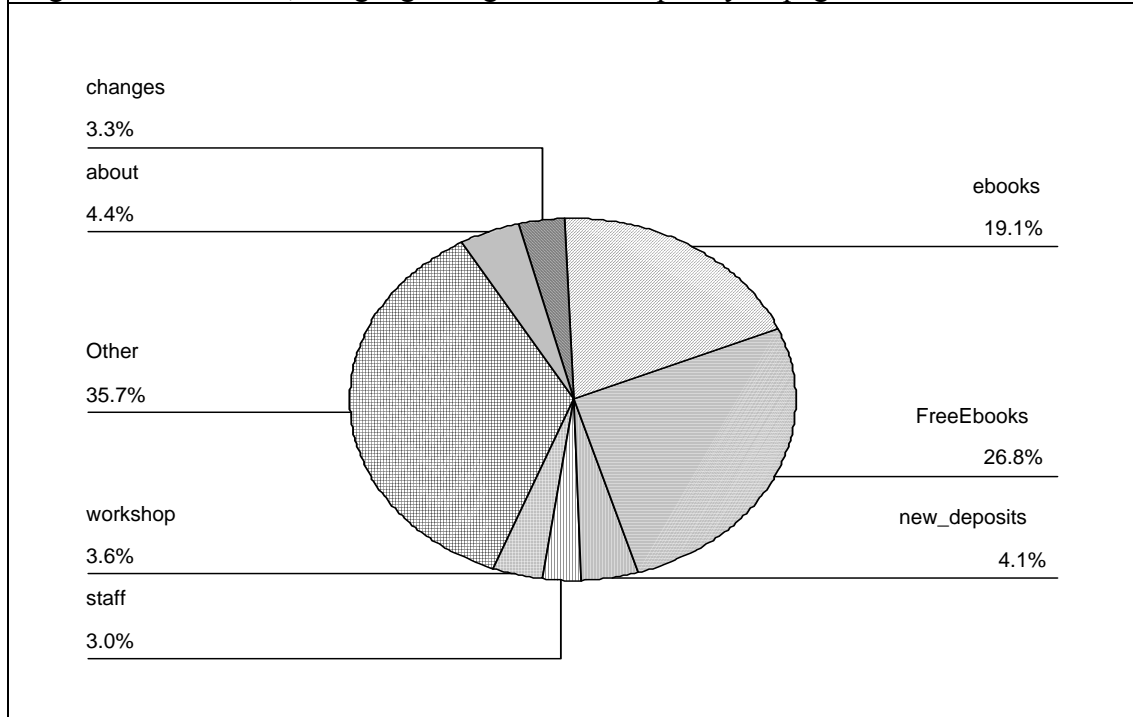
Figure 52: Performing arts – frequency of pages viewed

staff
5.6%

metadata-workshop
2.8%

designing-shakespear
8.3%

changes
7.8%

ahrb-advice
3.3%

Other
70.0%

advisory-committee
2.3%
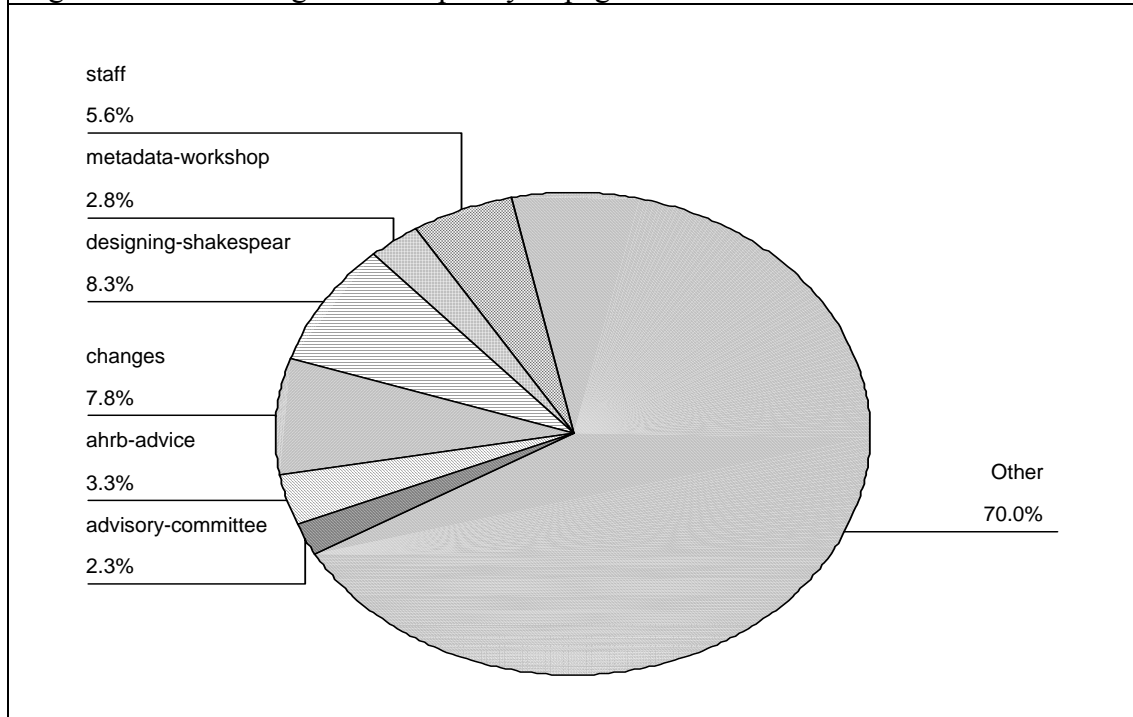
Figure 53: Archaeology – frequency of pages viewed

Other
61.7%

team
4.9%

stone
3.6%

project-management
4.1%

copyright-introducti
3.7%

deposit-formats
5.4%

evaluation
3.8%

guides
5.8%

invitation-to-deposi
7.1%
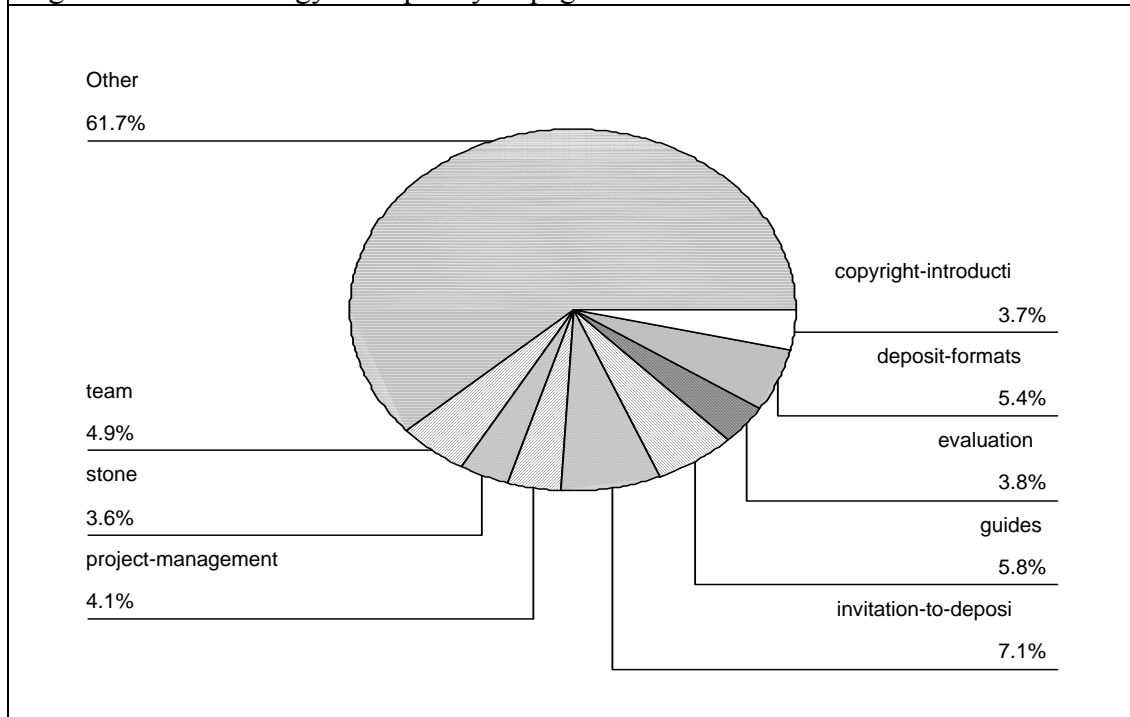
A5.4.5 AHDS Web-Log Analysis Conclusions

Site Usage

The AHDS site averaged 1-3,000 hits per weekday with an average of 600-900 sessions per day. These figures, however, are not an adequate measure of overall site visits since so many users accessed the AHDS at its individual sites rather than through its central server. The impact of commercial DNS access to the AHDS site is even more pronounced than for Humbul; but it is more accessed by UK users than Humbul.

Site Penetration

Academic users, as with Humbul, tended to make more 'serious' use of the site than those identified as coming from non-academic origins. Of the academic users, half visited 11 or more pages/views on the site during a visit. This suggests a satisfying depth of penetration to the site's resources.

Subject Distribution

In comparison with our overall Arts and Humanities research profile (**A2**), History and Archaeology significantly out-perform their cohort size. Visual Arts and Performing Arts perform in accordance with their profile. Languages, Literature and Linguistics under-perform in accordance with their profile.

A5.5 Artifact Web-Log Analysis

Despite repeated requests to do so, CIBER has not submitted its analysis of this material that has been supplied to them from this service provider. We shall submit an addendum to this report if it arrives.

A5.6 Individual AHDS Service-Provider Web-Log Analysis

Despite repeated requests to do so, CIBER has not submitted its analysis of this material that has been supplied to them from this service provider. We shall submit an addendum to this report if it arrives.

A5.7 Overall Conclusions

This analysis of the available web-log statistics provides a good deal of circumstantial detail about the traffic patterns of the service providers. But the conclusions that we can draw from it are disappointing modest and frustratingly inconsequential.

Although we have some indicative measures of **overall usage**, we cannot satisfactorily isolate academic and non-academic usage in the data. What is more, it cannot be reliably used in a comparative context. Many users accessed the AHDS individual sites rather than going through the central server. It is likely that the AHDS overall statistics significantly under-record its overall usage, whilst some traffic it records may well be internal to the service itself. The high volume of Oxford referrals in the Humbul statistics may also relate to traffic internal to its service. Any direct comparison of the impact of the AHDS as compared with the RDN subject centres is impossible on the basis of our evidence. The AHDS overall site-visits and session statistics may well have been higher than Humbul's if we include individual site traffic. Equally, the RDN subject centre traffics may be higher than that of the AHDS if we include the unknown usage statistics of Artifact. We have no way of knowing, such is the measure of our uncertainty about the reliability of the data to hand.

The indications of **subject distribution** provide some limited, but useful conclusions about the discipline-specific patterns of access to these services. In both instances, the most active users were from History – both in absolute terms and in comparison with their research cohort. In both instances, Languages, Literature and Linguistics were the least active users – in comparison with their research cohort. Philosophy, Law and Theology also appeared to be relatively inactive users. We hesitate, however, to draw

more specific conclusions about other subject domains from the rather fragmentary analysis furnished by the data.

The indications of **site penetration** reveal, unsurprisingly, that academic users tended to be more serious 'users' of the sites, both in terms of the numbers of pages/views visited and in terms of the amount of time spent upon the site. The statistics for those users that went on to link to an external resource directly from the Humbul site should perhaps be a cause for concern. On the other hand, we should probably place the statistics of AHDS site-usage alongside the downloads of collections from its sites [**A3, above**]. Taken together, this suggests that Arts and Humanities users in 2005 were finding more materials, and doing more with them, than they had done previously.

Anyone expecting to arrive at a picture of user-behaviour from web-log analysis is likely to be disappointed. It is a blunt instrument for analysis without complementary, detailed evidence of the user demographics in question. We therefore regard this evidence as being best adduced as part of a 'triangulation' approach, using it to confirm, strengthen or nuance, the conclusions we arrive at through our online questionnaire, focus groups, interviews, and Delphi analysis.

Appendix A6 Work-Package 3: First Focus Groups Report

WP3 Report prepared by Jared Bryson

SUMMARY

o Web use is ubiquitous and an integral part of a researcher's 'tool kit'
o Used primarily for accessing the increasing variety of primary and secondary sources specific to the researcher's needs
o Little use is made of WWW for any other purpose other than academic resources. Features such as web-based communication other than email were rarely referred to.
o Pushed news alerting for conferences and travel arrangements were used.
o Awareness of portals is mixed among researchers and rarely used when known.
o Google is preferred due to its comfort and ease of use and the volume of responses. However, the ability to validate and control the quality of search returns was considered a problem with most search engines.
o Institutional portals provided varying degrees of administrative control over non-research tasks.
o Controlling intellectual property and accessing the full array of literature were consistently raised as concerns across all disciplines

A6.1 AIMS AND METHODOLOGICAL APPROACH

The aim of the first set of focus groups was to capture qualitative data on the subjects indicated by the online questionnaire (**A4)** and the data log analysis (**A5**), a series of five focus groups and four one-to-one interviews were conducted among several of the AHRC subjects. The focus groups were drawn from the University of Sheffield departments of archaeology, history, biblical studies, music, and information studies. The interviews included scholars in the arts from DeMontfort University, including a lecturer in creative technology, music studies, digital imaging, visual arts and holographic applications.

Each of the focus groups consisted of three to seven participants from among the full-time contract researchers and lecturing staff. One of the focus groups did invite post-graduates, who were able to contribute their experiences of web-based work required as part of their research curriculum. All were asked a range of open ended questions designed to take advantage of group dynamics in a conversational environment (See annex). The number of questions ranged from eight to seventeen depending on the development of the responses and the need for prompts. Most focus groups lasted no more than an hour. The quotes appearing below are given timings in order to indicate the location within the overall sequence of the conversation. The selected quotations are meant to be representative of overall findings.

The questions put to the participants were intended to capture the broad contours of web and portal use among arts and humanities scholars. The intention was for these first series of questions to shape the second phase of focus groups primarily involved the presentation of screenshots demonstrating a range of features that might benefit researchers. The choice of the focus groups among the discrete subject areas was meant

to capture the disciplinary differences, picking up on the varying uses of the web for research and identifying any patterns in research vocabulary and culture.

In each focus group, the respondents were asked about their familiarity and use of the RDN portals and the AHDS sites respectively. Familiarity of the RDN portals was less in evidence than for the AHDS, though few members of the focus groups could recall precisely what the role of the AHDS was. No members of the focus groups actively used the RDN portals. No member of the focus groups had accessed a collection from the AHDS web-site.

A6.2 RESPONSES

A6.2.1 Purpose of Web-based Research and the Usefulness of Web Resources

Since the emergence of the World Wide Web as a ubiquitous tool of information access and communication researchers responded positively to it use. However, some frustration was admitted by some focus group participants particularly with the volume of information made available through search engines such as Google. Some respondents noted that their students found little difficulty adjusting to the rapidly changing web media, which may indicate a generation-based level of comfort with new technology. A few respondents considered themselves early adopters, especially those whose research focused on the use of the web such as information studies, music and biblical studies. However, the majority might be considered early or even late majority users of the technology within the framework first developed in Everett Rogers work on innovation (1995).

The first series of questions began by probing the researcher's web use habits, learning about what they use the web to do, how they did it and how well they thought it accomplished those tasks.

Responses to the following question often pointed to the near ubiquity of the web as a tool for researchers.

When you're conducting your research what do you generally use the web to help you to do?

> *Oh millions of things. Just about everything nowadays...* (Archaeology 00.23 minutes)

> *It's hard to reflect on how we use the internet because we… it's become something that we use all the time. It's become such an integral part of any work...*(Archaeology 10:30 minutes)

Three advantages of web-based research were frequently mentioned by all focus groups. First the **convenience** of accessing texts, images and artefacts within their particular subject specialty was a step-change over the need to travel to library collections—even if they were based at the researcher's own institution. The convenience of reviewing texts

from a laptop computer nearly anywhere, and the efficiency of searching those texts by key words has made the web and digital resources a boon for research.

> *It's used throughout the project. It makes all those things so much easier.* (Information Studies 5:30 minutes)
>
> *I think that there's something that emerges from the convenience of the Web. It's the lazy person's way of getting access to information resources, whereas otherwise we would have to get off our backsides and go to the library. What it does allow is that organic movement from subject to subject, following up hunches, chasing up ideas. There's a bit more free-wheeling component when you're using the web than when you're using paper-based resources* (Information Studies 1:30 minutes)
>
> *The library now has this access to NAXOS website so students can do less listening in the library and more listening at home online.* (Music 5:50 minutes)

The second advantage related to the **timeliness** of the information that could be accessed. The printed works found in institutional libraries were often burdened with a time-lag not found to the same extent among the digital resources.

> *The resources you'll find in the library are quite old, because of the print run time. So things may be more up to date if authors pre-disclose their papers online.* (Information Studies 2:00 minutes)
>
> *If it's in paper it's out of date* (Information Studies 9:00)

The third advantage was the **discovery** of new or otherwise unknown information. Several times during the sessions researchers referred to using the Web as a tool that helped them think—clarifying ideas and discovering new ways of approaching various research problems. While it was primarily used to search for specific resources that might yield specific answers, 'browsing as a way of thinking', and 'fishing trips' emerged from serendipitously 'surfing' through subjects and following interesting links. Others used the web to explore the 'state of the art', seeking out the publication and biographical details of peers within their community of research.

> *It's a way of thinking, isn't it? Just browsing is thinking.* (Information Studies 1:00 minutes)
>
> *Somebody had said that if the web had really been developed in Britain rather than in California, instead of surfing, we would 'potter'. And I actually do quite a lot of creative pottering around—following links and suddenly discovering something I never really knew this realm of discourse that might help with my research. But not quite knowing what I'm looking for, and hoping to stumble upon things.* (Biblical Studies 00.45 minutes)
>
> *Being old fashioned I look for things I already know that are there....it would be a combination of bibliographical resources and some collections online.* (History 00. 30 minutes)
>
> *I use it in two ways, one is "fishing trips": I don't know if the material is there but I'll want to see if there's anything; and then specific "validation exercises", where you know there's going to be some material but you want to find out if you've got the detail correct in your head or whether the publication data is correct or whether this person is still at X university or whatever* (History 00.56 minutes)

> *If it's academically validated it's really useful, but there's tons of erroneous rubbish out there, but then you've got to learn how to find your way around it.* (Music 4:53 minutes)

> *To find out what's already been written* (Information Studies 00.23 minutes)

Comments such as the following highlight the change in research culture. Previously one's time planning access to a library's resources occupied a substantial portion of a researcher's work, but currently, the **immediate access** to digital resources through the web has enabled work to be conducted far more rapidly and efficiently.

> *I've started to use online resources such as JSTOR and e-books. It does change the way you do research and allows immediate access to a big range of particular subjects* (History 3:55 minutes)

> *That's just extraordinary to have these printed resources available at your desktop at home where you might be working. It just changes the whole way I might be doing research (History 6:20* minutes*)*

> *It's the first place you look for anything. It's the easiest thing to do to just type it in where you're sitting. Even to stand up to go to the bookshelf takes a lot longer.* (Archaeology 1:23 Archaeology)

> *The ability to search made it so swift that I was able to do in a month what would have taken, oh, several months any other archive. And then you can archive it and go back to it. So I think to have primary source material online, and PDF-able would be fantastic. I know it's an enormous task but it's astounding what a fantastic research tool it really is, much much more powerful than I thought.* (Music 8:40 minutes)

Internet-based **communications** now facilitate instant dialogue with colleagues and specialist subject hubs even help to set up conferences.

> *I think sharing data is a huge thing. I mean the ability to share image data, images of anything. Images of microstructures or artefacts you can move those in huge volumes compared to what you used to, and that definitely enhances the quality of the work, the discuss-ability of the work* (Archaeology 10:20 minutes)

> *…their program allows you to take [conference] proposals and allows you to process them…you used to have to get all of the emails and send them around yourself. Now it's just all done from the website. You don't even have to say your proposal is accepted or rejected that's done automatically.* (Biblical Studies 3:20 minutes)

A6.2.2 Distinctive Research Practices within Subjects

Since the RePAH remit sought to establish trends in researcher's use of the Internet and web-based portals, there was a concern that the breadth of subjects in the humanities might prove problematic. Each subject area does possess a research culture with its own vocabulary and concerns. To identify where those boundaries might exist the focus groups were asked:

> *Can you think of any ways in which your discipline affects your use of the Web? In other words how does the fact that you're a/an (insert discipline) affect how you use the Web?*

Several of the individual researchers indicated that they might work across disciplinary boundaries, but collaborative work even among scholars within the same field was less common. However, there may be in interest in using the same types of data sets. It was as likely for archaeologist as it was for a lecturer in biblical studies to need access to geographical information system (GIS) data, maps or aerial photographs. Some general boundaries could be identified with regard to the emphasis placed on data among the five subject areas examined: musicians were concerned with audio data; archaeologists worked with three-dimensional artefacts; and information studies, historians and biblical studies dealt with texts (though information studies was likely to work with unprocessed data or be concerned with various organisation systems for all manner of information).

> *What we do overlaps with some many others: reading texts, history, politics, or whatever... It's an interdisciplinary exercise. Other than reading some texts in Greek or Hebrew I'm not sure there is anything we would do that others would not.* (Biblical Studies 6:40)

> *We bring together aspects of other, especially in archaeology, because it crosses disciplines, we bring together other aspects of things that people have done and then present them in a new way.* (Archaeology 26:40 minutes)

> *…EEBO and EECCO which is the Eighteenth-Century equivalent--it's quite extraordinary to have all these printed sources available at home…it changes the whole way I think about doing research.* (History 5:50)

> *[Information] is what we're studying and what we're studying with. Part of studying the Internet is studying communication…there are fundamentals to communication and the principles underpinning them go well before the Internet… the factors that motivate people to look for information were there before the Internet was around and they're still there. Those are also factors that probably contributed to the Internet becoming the success that it is. Whether to do with what it means to be human, psychology of humanity and ways of exchanging ideas, all those things are fundamental to humanity and not to the technology.* (Information Studies 9:30)

Several subjects blurred the boundaries between humanities and the social sciences. The use made of the web by various elements of society was not only an issue of interest for those in information studies but also among those disciplines such as biblical studies:

> *…There's a level at which the internet becomes one of the sources and the targets of the search in a certain kind of biblical studies.* (Biblical studies 9:30)

The role of 'resource discovery' as a primary task for scholars within the humanities appeared to be relatively ubiquitous across the humanities disciplines, with the value-added factors derived from the analysis and complex interpretation put upon the various texts and 'artefacts' retrieved in digital form. The musicians' use of digital audio file stood out from the others in terms of capacity and quality of data transfer. It was noted that as the quality of performances increased as technology advanced, and many within the field were helping to push the boundaries of the discipline.

> *A typical project for a student might be one gigabyte and that might be one project for one module. There are none of the bog standard systems that give them that kind of space. As the quality is going up and up and up, the sizes are going up and up…it's*

> *about 10 megabytes a stereo-minute...that's 4-8 gig in one project.* (Music 19:00 minutes)

> *There are a number of radio stations that do broadcast in 5.1 (surround sound) but it would be quite easy to get that streaming capability from the University website.* (Music 18:00 minutes)

It might be expected that any advances in the infrastructure that enabled these large, high-quality audio transfers would also be beneficial to scholars across the breadth of the Arts and Humanities.

A6.2.3 Awareness and Use of Portals

With the advent of powerful search engines such as Google, the role of the portal seems to have diminished considerably. The simplicity of Google's search field made it easy to use and therefore contributed to its nearly universal choice as an entry point to the Web.

> *I don't use them a lot, because I usually find things much faster if I Google, and I get straight on to what I need. And because I've done it a lot I see immediately from the little thing you get whether this would be relevant or not. Portals are a waste of time."* (Biblical Studies 13.5 minutes)

> *I've used them in the distant past, but not recently. Usually I just use Google and go straight to whatever I want.* (Archaeology 18:35 minutes)

> *For me the virtual environment that a portal offers...I don't like the self-contained, access to all but it's not really access to all. You know the easiest portal for me is Google.* (Music 24:54 minutes)

> *Google.* (History 7:30 minutes)

Another form of information distribution, the web log, was identified by a respondent in information studies as a type of portal. An acknowledgement of this and the characteristics of web logs used as portals may have ramifications for the development of more traditional portals.

> *A weblog can act as a portal…*(Information Studies 48:50 minutes)

The ubiquity and usefulness of search engines as a means to access the World Wide Web make portals seem old fashioned and thus somewhat confusing. We asked:

> *Do you ever use portals in your research?*

> *The definition of a portal is not so clear.* (Information Studies 15:20 minutes)

> *Not for research.* (History 7:50 minutes)

When portals are known and used the respondents suggested that their broad-based content was more useful in the initial, formative stages of research or when beginning a review of the literature on a general topic.

> *Good at the beginning of a search for the general overview of materials.* (History 13:50 minutes)

> *The systematic searching, the kind you do on Humbul is the kind you'd do at the beginning of a project, and I never seem to be at the beginning of a project. I'm always deeply into a project, and at that point you tend to think you know what's out there, you know what you need to know. You know where the material is. You may not <u>have had time to look at it all</u> but you know where it is, and so you don't feel that you want to that kind of comprehensive search for sources or bibliography that you might do at the beginning of a project. So it seems to me that a portal is most useful at the beginning of a project.* (History 13:40 minutes)

However, there was some discrepancy as to whether a portal offered specialist or generalist information, as seen in the following examples:

> *A lot of people use a search engine to find a specialist tool which is often a portal.* (Information Studies 15:10 minutes)

> *I've used British History online, but I don't know about the classic use of a portal because I've been targeting particular resources I know that are there.* (History 7:40 minutes)

The mediating role that portals play between extraneous and useful information was acknowledged by some participants.

> *I like the idea of portals, because diversity of information on the internet is sometimes too much. Portals help us to focus on a specific subject or a few related specific subjects. So portals give us a lot about a little, rather than a little about a lot. So when you are looking at portals you feel like you have a sort of control on the information. You don't feel lost. You feel like somebody else has already collected the relevant information so the job is much easier to you. And this feeling is very enjoyable for me.* (Information Studies 13:40 minutes)

> *It's more authoritative than typing something into Google. You need some system of peer review and quality control.* (Information Studies 23:48 minutes)

> *Hit overload. And you give up after about 20 reviews and you say this is an impossible task and I haven't got time to do more than this, I've got the sense of what there might be, but you haven't really done a systematic job.* (History 10:50 minutes)

For one user of the Humbul site even the mediating role was not enough to weed out superfluous materials:

> *In Humbul—where is it?! I actually had to go back to Google and look at the cached entry…Humbul had just rendered itself a page of links, and that's not a criticism, because it's still useful I suppose, but you're still overawed by the amount of stuff on the portal itself, and you have to Google just to get through the front page!...and when you think about what that portal should be doing it's just a bit crazy.* (Archaeology 32:00 minutes)

The awareness of portals for the Arts and Humanities researcher was mixed. If they were known their use was considered limited.

I like the idea of it but I don't use it. (Information Studies 34:50 minutes)

I don't know as much about them as I'd like to. (History 10:10 minutes)

Portals miss the point. (Information Studies 17:57 minutes)

It's funny the way that I've used Humbul is through Google, searching for something else. (Archaeology 17:10 minutes)

Portals for some reason miss the point…part of it is knowing that they exist…(Information Studies 18:20 minutes)

For respondents the mediating role of subject portals in selecting resources was considered too limiting. Several researchers worried that they felt like they might be missing something if they relied on a portal, especially when the information they needed was readily available through a series of searches on Google.

There is a sort of claustrophobia about portals. I never thought if them in this way before…there are loads of things I could find out better ways elsewhere so why bother…if I were looking for things in my research area I'd be just a bit worried that that portal is only as good as the people controlling it. If you're doing research then you should be at the cutting edge of whatever you're doing and you should be defining whatever it is you're using. It seems that they are just too self-contained. (Music 26:00 minutes)

Institutional portals were used primarily for administrative purposes, or for accessing the library catalogue and presenting teaching materials.

I suppose we use the university portal all the time… (Archaeology 18:15 minutes)

A6.2.4 Strengths and Weaknesses of Portals

Linked to issues of awareness are the strengths and weaknesses of portals. For the respondents the portal did not offer any clear advantages over search engines. The mediation could be seen as too controlling and the majority of specific resources were only the lower grade open access type. A prominent concern was over access to the entirety of literature within the field. The organisation of the subject portals were considered too primitive to be of much use and instead acted as a hindrance to their use. However, the ability to control the quality of initial web searches concerned many of the focus group participants. The trustworthiness of data was an issue, and yet the majority respondents were reluctant to use the mediated sources provided though the humanities subject portals.

Portals can be designed for specific research communities; however this can pose a problem for ease of use and accessibility.

It's just assumed that we know what we're doing…and they're not straightforward to use. (History 11: 40 minutes)

The quality of the materials available from portals was mentioned several times.

Portals such as Humbul catalogue open access materials, but what we've focussed on is those that that are only available on a subscription basis which by definition cannot be readily catalogued by these sorts of people because that's where the added value lies. So one of the problems with portals is that they tend to catalogue lower grade materials and the refined portals are those which are costing the institution something. (History 13:00 minutes)

One of the problems with portals is that they tend to catalogue low grade materials and the refined portals are those which are costing institutions. (History 13:20 minutes)

As real specialist I wonder whether a portal is good for us, because we're good at seeking out...we don't look at every aspect of archaeology...and we're good at seeking out the specialist information that we need, and filtering ourselves, so you'd need a hundred experts on the different aspects of archaeology to be there to provide the right stuff from a portal (Archaeology 12:45 minutes)

Since one of the value added advantages of portals is their role in mediating the vastness of the web, several participants wondered about the nature of the 'gate keepers' and the type of information being selected for viewing.

What a portal is doing is encouraging you to browse through a more restricted range of documents or whatever, which might reflect a more official view of what you should be reading on a subject. (Archaeology 22:30 minutes)

I think the important thing in a portal is who is the gatekeeper? Who selects what's authoritative? Are you as a researcher, willing to trust that person X in that nameless office that you can't see is selecting the authoritative sources that you can rely on? (Information Studies 25:30 minutes)

Portals were considered to lack an awareness of user search techniques. Unless the portal site took this into account it would be dismissed by researchers and simply not used.

The way people do research has not caught up with the possibilities of the Internet. (History 18:00 minutes)

You develop quite a feel...I think we all have an intuition in the way that we find this stuff, so that when someone from France says can you give me a bibliography on X, I'll sort of know that Humbul isn't going to help me on that, that it's just too specialised, it's just not going to work. Whereas I might get something from Historical Abstracts, and I'll go straight to that. Of course I might have been wrong, but that's the judgement call that one makes, and those judgement calls are sort of intuitive. So the first impression that you have of a site is very important, because it feeds almost immediately through to how you use it thereafter. (History 14:30 minutes)

The problem with things that are automated is that they cannot discriminate in any way the same as the human eye. So although you can to some extent find some useful stuff by having an automated crawler go and pull it in for you, it's going to pull in stuff that is totally inappropriate, and more the point stuff that is irrelevant. So you're going to suffer from information overload unless you can understand how to limit the stuff that comes in your direction (Information Studies 23:20 minutes)

> *I think a portal is a low-level activity. It's basically collecting things together that are useful, but it's actually not analysing anything and that's what research is about....knowing that those resources are there is part of your expertise.* (Information Studies 33:00 minutes)

Some participants wondered what advantage a portal offered that wasn't already being built into web browsers or desktop applications.

> *What does a portal do that my favourites list can't do?* (Information Studies 23:30 minutes)

However, in response one of the participants suggested that a portal was still useful for providing opportunities for new discoveries—the core task of researchers using the web.

> *I feel capable of distinguishing between good and bad things myself. A portal doesn't have to do that completely. If some of the things that it produces are relevant and I didn't see them before, that might be where it scores over your book mark. It might throw up things that you might not have thought that were there before.* (Information Studies 25:50 minutes)

One participant from the music focus group picked up on a weakness among institutional portals:

> *The university's bog standard package doesn't deal with a lot of the formats that we would use, so we have to find another way of doing things that almost by-passes the University's standards.* (Music 13:00 minutes)

> *They have gone too much down the marketing route without paying attention to the fact that the departments have requirements, specific requirements.* (Music 15:40 minutes)

The generic nature of institutional portals for this participant meant that the huge audio files needed for working digitally could not be accommodated and thus significantly hindered teaching and research.

A6.2.5 Desirable Features

If portals were to be improved or additional features put in place, respondents requested that the applications not simply duplicate existing features. Instead they should offer something new, be readily accessible, comfortable and easy to use. Portal features would certainly need to take into account their target users in order to avoid trying to be a 'jack-of-all-trades', and good at none.

> *You need to determine how deep and how broad, other wise your portal will be enormous. So you have to make a decision of who is the user. If there are specialists in an area they need deeper information, more specialised information. If there are intermediate users then they might need broader coverage.* (Information Studies 25:00 minutes)

The ability to search with a greater degree of confidence, control and with a finer granularity was also mentioned, as was the ability to search across multiple databases simultaneously and remember past searches. A single authentication feature was

mentioned by a few researchers concerned that accessing multiple publishers was difficult and could better be cleared through a single sign-in.

> *I would like to have a search function which I could customise so that I could tick do I want to search British History Online and Humbul and search a lot of different sources…I find that I'm going to different sources and conducting the same search, which is very time consuming, and then you got the job of homologating all the results…if Athens could be placed on the institutional portal, the MUSE portal, and then you could search within all those resources.* (History 26:30 minutes)

> *Very elementary search engines don't allow you to be very specific about what you're searching for. It seems that web designs are moving to a simpler and simpler search box, when in fact what you need…when you want to, you need to be able to be a lot more specific about what you're looking for.* (History 11:20 minutes)

> *It's not that Google has too much, it's a matter of how it's prioritised… I get the sense that it's more for general things and it's the specific things that are a problem.* (Archaeology 19:30 minutes)

Researchers used the web primarily to access literature within their community; therefore access to the entire breath of literature was a frequent request.

> *The first thing it needs is access to every journal in that subject Because if you go to a portal and you think that you're missing out just by using that you're just going to go back to Google—because it's so powerful you can get most stuff. I don't know what the purpose of filtering is, unless you already get everything and it saves you a step.* (Archaeology 27:50 minutes)

> *Research momentum stops there and then…it ruins your organisation for the day. If they could set up the finances somehow so that…couldn't they just have access to all journals and have it on a pay per view? …So that the research process is never stifled. And for the sake of humanity that's a very good thing to do and if it creates more research, more intellectual endeavour they are morally obliged to do. And portals could allow that if the pay to view thing was worked out.* (Archaeology 29:10 minutes)

Pushed alerts, perhaps via RSS news feeds was a notable feature. This was to afford access to the latest additions to bibliographical databases, notification of conferences and calls for papers, and most importantly news of funding announcements.

> *News I suppose…professional newsgroups…in terms of conferences and a useful review section* (History 19:00 minutes)

> *Calls for paper and calls for research funding and pushes it my way…instead of me having to trawl through all the funding bodies all the time.* (Information Studies 27:00 minutes)

> *News feeds stimulates your interest to look at other things.* (Archaeology 30:30 minutes)

Some measure of control over quality was mentioned on several occasions. The volume of returns form search engines such as Google, and the validity of those results needed to have some means of being controlled or filtered.

> *For me in research terms I want a portal to be credible and rigorous, and specialised, because that's what you're taught to believe when you're using these things. It's more*

> *authoritative than typing it into Google. You need some system of peer review and quality control if you're using it terms of research.* (Information Studies 24:00 minutes)

It was also suggested that the returns that are valuable to a researcher should be able to be catalogued and bookmarked in a more sophisticated way. Since web sites are sometimes not sustainable over long periods, moving the discovered data to one's bookmarks or desktop simple shifts the problem of storing and indexing from the web to the individual's own computer. The ability to more centrally and sustainably access information was also raised as an issue.

> *I'd like to be able to bookmark resources or link part of a resource…so you wouldn't have to fish it out from that original source…I'm finding more stuff downloaded to the hard disk, which is transferring a problem from one big domain to another big domain. And it's a big domain which has even more rudimentary search facilities than is available on the internet.* (History 30:30 minutes)

Since the primary use researchers put to the web was accessing the literature within their field most wanted even more depth and breadth of digitised texts. For the subject specialists such as archaeologists and musician working with large image or sound files the ability to move large datasets, such as maps or music performances was also a concern.

The web was understood to be increasingly useful for professional networking and sharing information with colleagues, whether through official peer-reviewed journals online or more informal posts on weblogs.

> *A portal which directed you at that sort of living conversation (peer-review-type open for comment) is probably more useful to than a researcher…*(Information Studies 42:10)

> *Evaluating stuff…You get a request to evaluate a research proposal and you tend to go increasingly on to the web to see what the context is, just to have a feel for how robust this proposal is.* (History 19:30 minutes)

> *Just to find out about people in the profession.* (History 21:20 minutes)

One other issue that some focus group participants raise centred on aspects of intellectual property and ownership of web-based resources. Tools that facilitated the process of securing permissions we welcomed.

> *Copyright is a major deterrent to academic freedom.* (Biblical Studies 28:00 minutes)

A6.2.6 Politics and Funding Issues

The tenor of most focus group respondents was negative about an additional portal tool if it held the possibility of duplicating any existing web or software applications. Most were worried that funding for such a programme could be better spent elsewhere.

> *I'm not sure that pumping a lot of money into this will really help because businesses already see their own interests in funding this.* (Biblical Studies 27:00 minutes)

The funding that they give is limited to a certain amount of money over certain periods so they can't really improve any projects their funding because they're not renewable. (Biblical Studies 35:00 minutes)

Finances are tight for just about everything, so unless there were funding forthcoming, I suspect it would not be high on the agenda. (Information Studies 28:20 minutes)

I don't want to pay any more money for something that I already have access to that makes it easier. To be perfectly honest I find that quite offensive. What I don't want to do is give in to people who are a bit lazy, and yes it might be difficult, but that's part of learning. You can't expect everything to just be there. It's not about saving time, it's actually about providing something new and useful. So what I would like to see anything you're going to spend money on giving us something new and not an interface to something we've already got. It would be about new subscriptions…if we've got new sources of information then I think that's the only thing we should be spending money on…equipment and information. (Archaeology 33:20 minutes)

Raising the profile of the existing services would be the best way to see an increase in portal usage. As seen above, awareness was relatively low and use even lower. This was due primarily to the more powerful and useful tool offered by Google and other web search engines.

…they need to make a much more rigorous effort to get them out there, if they want them to be more widely used. And that's the only way they'll know if they're truly valuable or not. I think it is an issue of awareness in the first instance…because I still use the ones I was told about…(Information Studies 47:40 minutes)

Additional or improved features should emerge from the researchers' needs rather than being developed simply because they can. Access to the largest number of high-quality texts and artefacts that drive research was given as a priority by most focus group participants. Therefore, portal tools should secure greater access to journals and other outlets of peer-reviewed research, and simultaneously preserve intellectual property rights and satisfy copyright demands would better serve existing researcher's web use.

Any effort by portal providers that were simply making use of technology for technology's sake when other issues were a priority would be frowned upon by all disciplines participating in the focus groups. However, there was also awareness that many of the efforts to improve service might not be readily apparent or have measurable outcomes.

A lot of sites (departmental resources) are there just as an intention of good will…ticking boxes. So ticking-box portals…there's a lot of that stuff around. Everybody doing useful links pages which they haven't really put much thought into. (History 38:30 minutes)

When you spend money on a portal you cannot get any money back. It's very useful to bring more control over the chaotic situation on the internet…but if a company or a department spend money how can you see the result? (Information Studies 29:30 minutes)

A6.3 CONCLUSIONS

The web is now a ubiquitous tool for researchers within the arts and humanities, used primarily for accessing a growing corpus of digital texts, images, audio and video resources. However, portals do not play an important or useful role among respondents in this series of focus groups. Most were satisfied with their existing applications, though would prefer greater access to their subject's literature. The Google search engine was the preferred application for accessing the Web, primarily for its simplicity and ease of use. However, the quality and quantity of search returns was a serious problem. Therefore, what was wanted were even greater refinements in searching and controlling Web information. For those in the music and archaeology focus groups, the size of storage and the ability to move large files through email or across the Net was a serious infrastructural concern, as was the institutional support for ICT related issues.

The issue of copyright and the issue of intellectual property rights were also seen as a growing concern. With greater and greater access being given to resources, the need to use materials for teaching or research meant that proprietary control was a major concern for researchers. Any technologies that might facilitate use of these resources should be a priority.

Since customisation, choice and the personal ability to control access to resources were vital qualities for respondents, and new tool that failed to offer these features would be likely to fail and find little return for investment.

Research Portals in Arts and Humanities (RePAH) Focus Group Questions (FIGURE 1)

Discipline: _____Group Size: _____

Date: _____Location: _____

Moderator: _____Recorder: _____

Question Category	Guide Questions & Prompts
	1. When you're conducting your research, what do you generally use the web to help you to do? *PROMPT: list of possible tasks from e-questionnaire*
Usefulness	2. How well does the web help you to achieve that?
Distinctiveness of the Discipline	3. Can you think of any ways in which your discipline affects your use of the Web? In other words, how does the fact that you're a/an (insert discipline) affect how you use the Web?

Portal Awareness	4. Do you ever use portals in your research?
	   a. ⟋ Which one(s) are used for your field of research?
	   b. How frequently?
	   c. If you don't use portals, why not? What might encourage you to use them?

Portal Usefulness
 	
Portal Strengths
Portal weaknesses
	5. What do you think about the usefulness of these portals?
	   a. What do you <u>like</u> about the portals that you use?
	   b. What do you <u>not like</u> about portals?
	PROMPT: define a portal and give an example for the subject's discipline—list from Humbul/Artifact?

Desirable tools & Research Needs	6. If you could have a web-tool that could assist you in your research, what features would it have? a. What information do you need to do your research that could be accessed electronically? *PROMPT: give an example of various tools that could assist various aspects of the research cycle*

Politics & Funding	7. How would this department be able to support a portal-based web tool?
	   a. Financially—if you had to subscribe to a service
	   b. Pedagogically—if you had to learn a new skill
	   c. Politically—if you needed the service in the face of Administrative reluctance
	ALTERNATIVELY: What would this tool have to offer in order to be supported by this Department?

Contribution (gets at capabilities of support A&H informatics)	8. What could your department contribute to a portal service?

Appendix A7 Work-Package 4: Analysis of the Delphi Exercise

WP4 Report prepared by Robert Ross

A 7.1 Introduction to the Delphi exercise

The Delphi technique is a systematic, iterative predictive research method based on independent inputs from a panel of experts. It measures the degree of consensus among the panel regarding future events where the decisive factors are subjective, and not knowledge-based. Delphi was developed by the RAND Corporation in the late 1960's. The technique reaps the benefits of group decision making while insulating the process from the limitations of group or peer pressure and overly dominant individuals.

The technique involves iterative rounds of questionnaires where responses are re-circulated so individuals can reconsider their opinions in the light of the responses of the panel as a whole. In face-to-face discussions or focus groups a dominant personality may exert much greater influence than their expertise should allow, this technique avoids that risk. Within RePAH the exercise entailed asking arts and humanities research practitioners what ICT tools or services they considered should be available in the future to support their research. For the purposes of this exercise practising researchers were regarded as experts in that they are highly knowledgeable about their own research processes, those in their particular domain and about research methods generally.

The original timescale for the Delphi exercise extended over the period from mid September 2005 to mid January 2006 and was to be conducted via the website. Timescales were revised to take account of delays at the start of the project. The revised timescale was as follows:

	Task Name	Duration	Start	Finish
1	Preparation	5 days	Mon 13/02/06	Fri 17/02/06
2	Circulate questionnaire 1	5 days	Mon 20/02/06	Fri 24/02/06
3	Analyse results	5 days	Mon 27/02/06	Fri 03/03/06
4	Circulate questionnaire 2	5 days	Mon 06/03/06	Fri 10/03/06
5	Analyse results	5 days	Mon 13/03/06	Fri 17/03/06
6	Circulate questionnaire 3	5 days	Mon 20/03/06	Fri 24/03/06
7	Analyse results	5 days	Mon 27/03/06	Fri 31/03/06

Figure 1 Timescale

A7.2 Sample

The sample comprised all members of the focus groups plus those respondents to the online survey questionnaire that had agreed to being contacted for further information and known experts from researchers similarly identified through the Aria project (n=109). An 'RSVP' email was sent to this combined list generated in late February explaining the aims, objectives and what would be expected of participants. An opportunity to withdraw from the exercise was provided on contacting the project manager directly by email. Three of the participants decided not become involved with the exercise leaving a total of 106 within the sample.

A7.3 Functions

The list of functions used in the exercise was identified via the focus group interviews that took place in **WP1 (A4, above).** These were: aggregation of data for searching and analysis, quality control and ranking system, online collaboration tools, grid connection / services, personalisation and bookmarking, desktop video conferencing, peer review facility, pushed alerts for conferences / papers / funding, access to all journals and finally copyright management. An explanation was provided to ensure all participants understood what was meant by each term. The functions were to be scored as being invaluable, quite important, not very important or irrelevant to the participant's future work.

A7.4   First round

In the first round there were 21 respondents to the exercise, a response rate of 21%, and the ranking scores of the functions was as follows:

	Invaluable	Q. important	Not v. imp	Irrelevant
Aggregation of data	8	7	4	2
Quality control	8	8	3	2
Online collaboration	1	10	5	5
Grid connection	2	7	8	4
Personalisation/bookmarks	4	13	4	0
Desktop video conf	0	7	12	2
Peer review	3	12	6	0
Pushed alerts	12	4	5	0
Access to all journals	20	1	0	0
Copyright management	6	11	3	1

Figure 2 Delphi First Round

Using the data produced the following graph
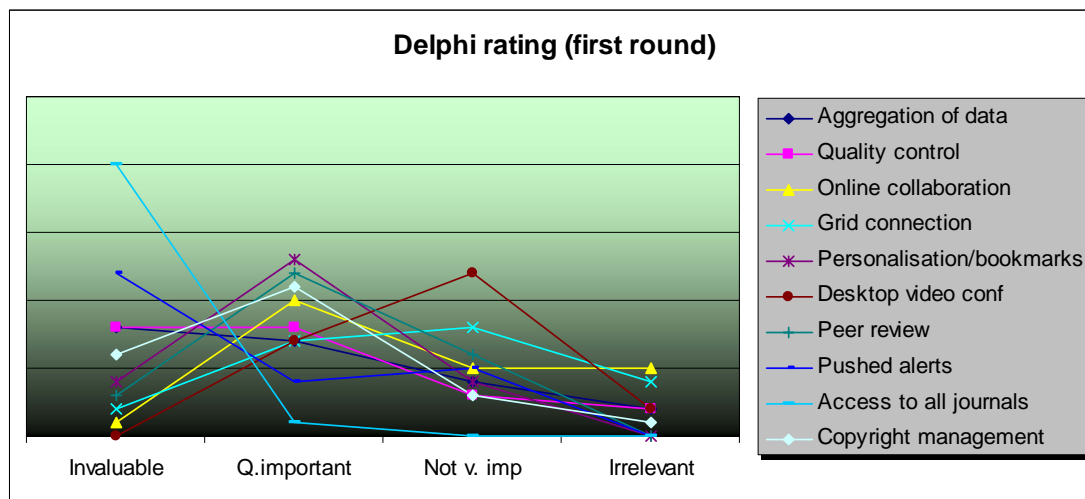
Figure 3 Delphi Rating (First Round)

Using the initial scoring of 'Invaluable' it can be seen that Access to all journals was voted the most important feature, followed by Pushed alerts. The next features to score the most 'Invaluable' votes are (jointly) Aggregation of data and Quality control. In order to differentiate joint 'Invaluable' scores a weighting system was used. Each score for 'Invaluable' was given a weighting value of 4, and that of 'Q. important' a weight of 3. The data obtained from using this method of calculation enabled a more granular ranking and placed the function Quality control above that of Aggregation of data. No further joint scores were present, so ranking reverted back to the most numbers of votes within the 'Invaluable' category. This method of calculation was used for all following joint scoring to enable a ranked listing to be produced.

A number of emails were received pertaining to this initial stage of the exercise. They fell into three groups:
1. those wishing to be removed from the list (3)
2. those stating they would be happy to be involved in the future rounds (7)
3. those wishing to receive details of the outcome from the exercise (2)

Using the data a new list was drawn up ranking the features according to their score, producing the following newly ranked list:

1. Access to all journals
2. Pushed alerts
3. Quality control
4. Aggregation of data
5. Copyright management
6. Personalisation/bookmarks
7. Peer review
8. Grid connection
9. Online collaboration
10. Desktop video conferencing

A7. 5 Second round

A second round was initiated and the respondents informed of the newly ranked list. They were requested to re-score the list based on their response to the 'community' perception of what was deemed more or less important.

Initially there were only 3 responses to this stage of the exercise, but a follow-up email prompted more involvement from the list. At closure of the round there were 18 respondents in total, a 19% response rate, and the following data was obtained:

	Invaluable	Q. important	Not v. imp	Irrelevant
Access to all journals	8	4	0	0
Pushed alerts	3	4	4	1
Quality control	2	5	4	1
Aggregation of data	1	7	3	1
Copyright management	0	4	6	2
Personalisation/bookmarks	2	5	3	2
Peer review	2	3	5	2
Grid connection	0	2	8	2
Online collaboration	0	2	7	3
Desktop video conferencing	0	1	3	8

Figure 4 Second Round

From this data the following graph was obtained:
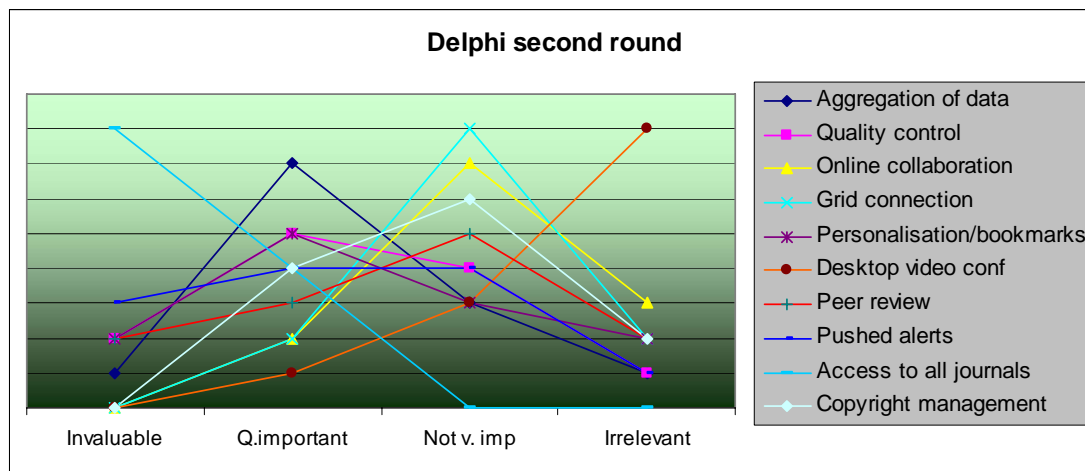
Figure 5 Delphi Rating (Second Round)

The profile of the graph is much less smooth than the first round, but this may be partially explained by the lesser numbers of respondents participating in this round. The lesser number means that each vote carries more weight overall and can cause a higher degree of variance across the data.

It can be seen from the data that Access to all journals and Pushed alerts are again the most highly rated features. The next three features are jointly scored and so the weighting calculation was used to differentiate them. The last four features also all scored jointly in the 'Invaluable' category. However, when the weighting calculation was used another joint score was produced. In order to calculate further granularity a second level of weighting was introduced which consisted of allocating all scores from the 'Not v. imp' category a 2. This enabled the features to be ranked and produced the following list shown here with the list from round 1 for comparison:

List from round 2	List from round 1
1. Access to all journals	1. Access to all journals
2. Pushed alerts	2. Pushed alerts
3. Quality control	3. Quality control
4. Personalisation/bookmarks	4. Aggregation of data
5. Peer review	5. Copyright management
6. Aggregation of data	6. Personalisation/bookmarks
7. Copyright management	7. Peer review
8. Grid connection	8. Grid connection
9. Online collaboration	9. Online collaboration
10. Desktop video conferencing	10. Desktop video conferencing

Figure 6 List from Rounds 1 and 2

It can be seen from these results that the features ranked mostly 'Invaluable' and mostly 'Irrelevant' have not changed. It is only the middle ranked features that have changed position relative to each other.

In this round the facility to comment on the choice of features was provided, and produced a number of qualitative data, some concerning the feature's relevance to the respondent and their research:

> *"I hope your research will bear in mind that there is a penumbra of researchers who are NOT in full time (or even part time) education, like myself, and have to make do with whatever they can acquire access to?"*

> *"I'm sure you've thought of this, but circumstances and the changing needs of different research projects, will mean changes in the importance of these features."*

Others related directly to the exercise itself:

> *"I remember that I had assessed no. 1 and 10 in the same way last time. No. 2 is not useful for me, so I'm sticking to my low ranking."*

> *"I probably changed my mind a little to reach an agreement within a perceived "group"."*

The last two comments highlight the process inherent within the Delphi exercise concerning group consensus, and that the respondents are aware of this and have reacted accordingly.

A7.6   Third round

The third and final round was initiated and the list informed of the newly ranked list based on data from the previous round. They were again requested to re-score the list based on their response to the new 'community' perception.

At closure of this final round there were 27 responses, a response rate of 28%. It was noted that there were far more respondents in this round and this may have a bearing

upon the outcome as obviously not all respondents took part in every round. This can be attributed to the large list of contacts and the anonymity allowed to the respondents; identification of those who took part in any single round was not possible and therefore filtering of responses to those who had previously taken part was not feasible. However, if the premise of the Delphi exercise is that community consensus will produce the best results, then the higher numbers within this final round can only serve to identify the most relevant features.

The following ranking scores were obtained from the final round:

	Invaluable	Q. important	Not v. imp	Irrelevant
Access to all journals	10	1	1	0
Pushed alerts	2	8	2	0
Quality control	5	3	1	3
Personalisation/bookmarks	1	9	1	1
Peer review	3	4	5	0
Aggregation of data	2	8	0	2
Copyright management	1	8	2	1
Grid connection	2	3	6	1
Online collaboration	1	4	6	1
Desktop video conferencing	0	2	4	6

Figure 7 Final Round

From this data the following graph was obtained:
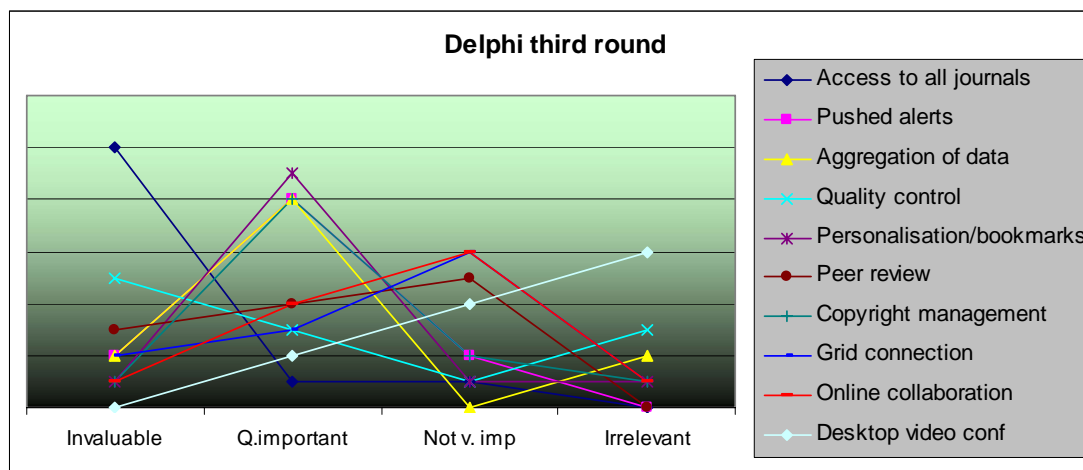
Figure 8 Delphi Rating (Final Round)

The profile of the graph is not as complex as the second graph, but is also still not as smooth as the first. This cannot be attributed to the lack of numbers as there were more respondents participating in this round than any other. However, looking at the scoring there were more joint ranking in this round than any other, which is not immediately apparent from the graph. Weighting had to be used twice and secondary weighting once to determine the features ranking.

The following is a list showing all ranking from each round.

List from round 3	List from round 2	List from round 1
1. Access to all journals	1. Access to all journals	1. Access to all journals
2. Quality control	2. Pushed alerts	2. Pushed alerts
3. Peer review	3. Quality control	3. Quality control
4. Pushed alerts	4. Personalisation/ bookmarks	4. Aggregation of data
5. Aggregation of data	5. Peer review	5. Copyright management
6. Grid connection	6. Aggregation of data	6. Personalisation/ bookmarks
7. Personalisation/ bookmarks	7. Copyright management	7. Peer review
8. Copyright management	8. Grid connection	8. Grid connection
9. Online collaboration	9. Online collaboration	9. Online collaboration
10. Desktop video conferencing	10. Desktop video conferencing	10. Desktop video conferencing

Figure 9 List from Rounds 1, 2 and 3

Access to all journals is again the most highly rated feature. Peer review appears to have taken on more significance as the rounds progressed whilst Pushed alerts has been relegated to a lower level of importance.

Although there was the functionality available to comment on the scoring or ranking of the features in the exercise, no respondents used it to express any further opinions.

A7.7 Conclusions

This Delphi exercise was aimed at creating group consensus on a list of possible functions available in a portal. This was achieved by using anonymous data to create an iterative ranking list of these functions so that personal standing would have no bearing on the outcome. Free text responses confirmed that this aspect of the exercise was clearly understood by the participants. However, allowing anonymous responses meant that there was no way of checking whom had or had not taken part in each round, and therefore it is not certain that the same people responded each time. In fact, this can be seen to be the case as the number of responses in the final round was more than in any previous. This does not mean there was no consistency within the sample replies, as the free text responses confirm that some of the respondents took part in at least two rounds. A classic Delphi exercise uses an identified small sample of experts (6 – 10) thereby ensuring that this discrepancy does not occur. The methodology used by the project in this exercise did not follow this procedure, as responses from a wide diversity of user were required to help identify those functions that are most useful to the community at large. It was seen that a fluctuation in participant numbers was an acceptable risk to ensure wide community engagement.

After three rounds of the Delphi exercise only three features remained in their original positions: Access to all journals was always rated the most important, whilst Online collaboration and Desktop video conferencing were rated the least important. Although there appears to be significant movement within the middle ranking functions, it is not as simple as it appears. In the first round there were only two functions ranked equally. These had to be weighted to obtain a rank result. In the second and third rounds there were two sets of functions equally weighted consisting of at least three individual functions in each set, some of which had to go to a second level of weighting to obtain a rank position. Taking this into consideration, the rank position of each function cannot be given too much importance and only a general inference as to their meaning can be made. What does seem to appear as an overall pattern, is those functions relating to individual activities attracted higher ratings as the rounds progressed. This can be seen with the movement of collaborative functions such as Peer review and Grid connection to the bottom end of the ranking. while functions based on individual effort such as Aggregation of data and Copyright management, moved up the rankings.

Similarities between these findings the online questionnaire responses and statements made by the focus groups suggest that the exercise provided a valuable insight into the needs and wants of a wide selection of the current arts and humanities community, that is confirmed by these other sources.

A7.8 Covering letter to sample population

Dear xxx
You have already helped the RePAH project to understand how ICT tools and services are being used currently by arts and humanities researchers and to identify other kinds of features that the research community might find useful. This valuable information has enabled us to begin to develop a picture of the kinds of services that could be made available in future and we would like to invite you to help us to shape that future by helping us to prioritise these ideas and to make sure we haven't overlooked any important functions. Would you be willing to take part in a short online exercise with a small number of other participants? The exercise entails considering a list of x functions and rank ordering them in terms of their importance to your research. Your responses will be pooled with those of other respondents and re-circulated in two further rounds to allow you to reconsider your opinions in the light of the responses of the panel as a whole.

A7.9   Text within the first exercise

RePAH Delphi Exercise
You have already been helpful in enabling the RePAH project to understand how ICT tools and services are currently being used by arts and humanities researchers. We now need to develop a picture of the kinds of services that could be provided in future and would like you to prioritise the functions listed in order of importance to research in your domain.

Please rate the following features:

	Invaluable	Q. important	Not v. imp	Irrelevant
Aggregation of data				
Quality control				
Online collaboration				
Grid connection				
Personalisation/bookmarks				
Desktop video conf				
Peer review				
Pushed alerts				
Access to all journals				
Copyright management				

Aggregation of data for searching and analysis: Accessing databases from multiple locations simultaneously, then bringing useful data together into one place for analysis and presentation. Data in this instance can be composed of digitised text, images, audio or video.
Quality control and ranking system: Searches would yield web sites and journal articles with grades of reliability based on a universal standard of validation, setting the search against a list of all potential hits with reasons for not including them in the validated list.
Online collaboration tools: Enabling work to be done on the same set of data (or even multiple sets of data) by more than one researcher, even if they are in different locations.
Grid connection/services: Internet - enabled collaboration between researchers, from different institutions, that typically involves secure access to distributed data, computing power and software.
Personalisation & Bookmarking: Automatic notification of any copyright information and use restrictions associated with a file when you access or download it and offering payment options at the point of use.
Desktop video conferencing: Using one's personal computer to conduct high-speed, high quality conversations over the WWW, rather than needing to access specialised facilities.
Peer review facility: The feature enables the data user to participate in the peer review process with anonymity and within the administrative criteria established for each particular subject specialty.

Pushed alerts for funding/conferences/papers: This feature picks up funding alerts from various sources, including research councils, government agencies, private foundations and international organisations. The same alerting service provides regular notification of conferences, calls for papers and new publications in the researcher's field of interest.

Access to all journals: Access to an array of primary and secondary literature, some of which may not be taken by a university library, but are nevertheless necessary and specific to a researcher's subject specialty. The portal provides access to journals including those discovered serendipitously and held by commercial, subscription services.

Copyright management: Automatic notification of copyright access and use of specific images, texts, audio and video downloads, offering permissions or royalty information.

Thank you for taking our survey. Your responses will be pooled with other respondents and re-circulated in two further rounds to allow you to reconsider your opinions in the light of the responses from the panel as a whole.

Appendix A8 Work-Package 5: Managed Research Environment Demonstrator

WP8 Prepared by David Gerrard and Stephen Brown

A8.1 Introduction

This is the report on work package 5 of the RePAH project. For the full report see http://repah.dmu.ac.uk/report. The main thrust of the RePAH investigation has been evaluative, that is to say it aimed to "discern patterns of use and to collect qualitative statements regarding the use and improvement of the various [….] components."[2] In broad terms this approach can be situated within the design-based research paradigm (Barab and Krishner 2001; Brown 1992; Collins 1992; Sandoval and Bell 2004; Shavelson et al. 2003). Design-based research is carried out in a continuing cycle of design, enactment, analysis and redesign (Design-Based Research Collective 2003). Within this study we have picked up the cycle at the enactment stage, conducted an analysis of the current picture and used the redesign stage to explore user reactions to possible future functionality through prototype demonstrators.

The purpose of the demonstrators is to obtain formative evaluation feedback that can guide further development. User feedback can also help developers to improve their understanding of the problem being tackled and help potential users to refine their own understanding of their preferences and needs.[3] This is a vital step in the design and development of all products because even if the resulting product is not perfect, it will be better than if no user testing were carried out at all. It is never too early to test and involve future users in the design process. Ross et al[4] recommend using methods such as focus group discussions and interviews for this kind of evaluation. The RePAH project therefore used a second round of focus groups and interviews to present a series of examples to researchers to gauge their responses to the functionality offered. The examples were based on the lists of features generated by the questionnaire (Appendix 4), combined with the outcomes of the interviews and focus groups (Appendix A6) and subsequently refined by a Delphi exercise (Appendix A7).

From this list, a series of wireframe graphical mock-ups was created to be evaluated in the final set of focus groups, in order to judge the reaction of researchers and to elicit further requirements for research portals.

The shortlist of requirements was:

1. Ability to conduct simple searches across disparate data collections.
2. Ability to share ongoing research work, notes and ideas with research collaborators.

---

[2] Peterson, E., York, V. 2003 User-Evaluation of the Montana Natural Resource Information System (NRIS). D-Lib Magazine July/August 2003 9 (7/8)
http://www.dlib.org/dlib/luly03/peterson/07peterson.html
[3] Ross, S., Anderson, I., Green, D., Albrecht, K. DATE The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials. Humanities Advanced Technology and Information Institute, University of Glasgow. http://www.nyu.edu/its/humanities/ninchguide/XII/
Consulted 06/02/02
[4] Ibid.

3. Ability to publicise and disseminate completed work, and comment upon other such work completed by peers.
4. Ability for comments / reviews / peer moderation to influence searches by flagging up content that has been deemed legitimate.
5. Ability to browse through disparate resources as well as search.
6. Moderation, submission and creation of content by community as opposed to central authority.
7. Inclusion of news feeds and current event information.
8. Ability to create new searches within the context of existing searches.
9. Inclusion of information background information about the creator of a piece of content, which would allow the user to assess their "point of view".
10. Inclusion of IPR and copyright information about resources.
11. Tracking of the user's use of resources discovered via the portal.

NB the requirement to access all journals was not explicitly included since journals are content, whereas the demonstrator was primarily concerned with functionality. Journal access is subsumed within requirements 5, 6 and 8 above.

A8.2 The Demonstrator

The demonstrators have been designed to be essentially modular in nature to allow extension and personalisation. As a result, they do not cover all the potential functionality of a system of this sort. Instead, the following are highlighted:

- **The system homepage:** what the researcher would see when they logged on using Shibboleth or similar authentication system.
- A typical **set of search results** that the user would see after conducting a Google Scholar search from within the system framework.
- An example of an **annotated web page** that a researcher has visited and provided comments about.
- An example of the **usage history** for a resource: in this case a paper in an online repository, though it could be a website, an online article, an entire journal, a dataset, a book from the library etc.
- The researcher's **bookmark** management system. Again, all types of resources could be bookmarked, not just web pages.
- The researcher's **online CV**. This would contain a short biography, their current job title and location and information about their projects (current and previous), their professional associations and a record of their publications.
- A **project management** page showing details of the project team and linking to all shared documents generated by the project, as well as email and shared bookmarks that team members had collected.
- A list of the researcher's collaborators or **research partners**. This page would also provide access to all the documents shared by research partners, all the email sent by and to them, and all the bookmarks they have shared, as well as links to their online CVs.

The wireframes and more in-depth notes regarding them are contained on the pages that follow. The diagram below shows a high-level architectural diagram showing how the demonstrator system might work; involving collaboration between software on the user's desktop, servers hosted by individual institutions and a centrally managed (presumably by the JISC) database.
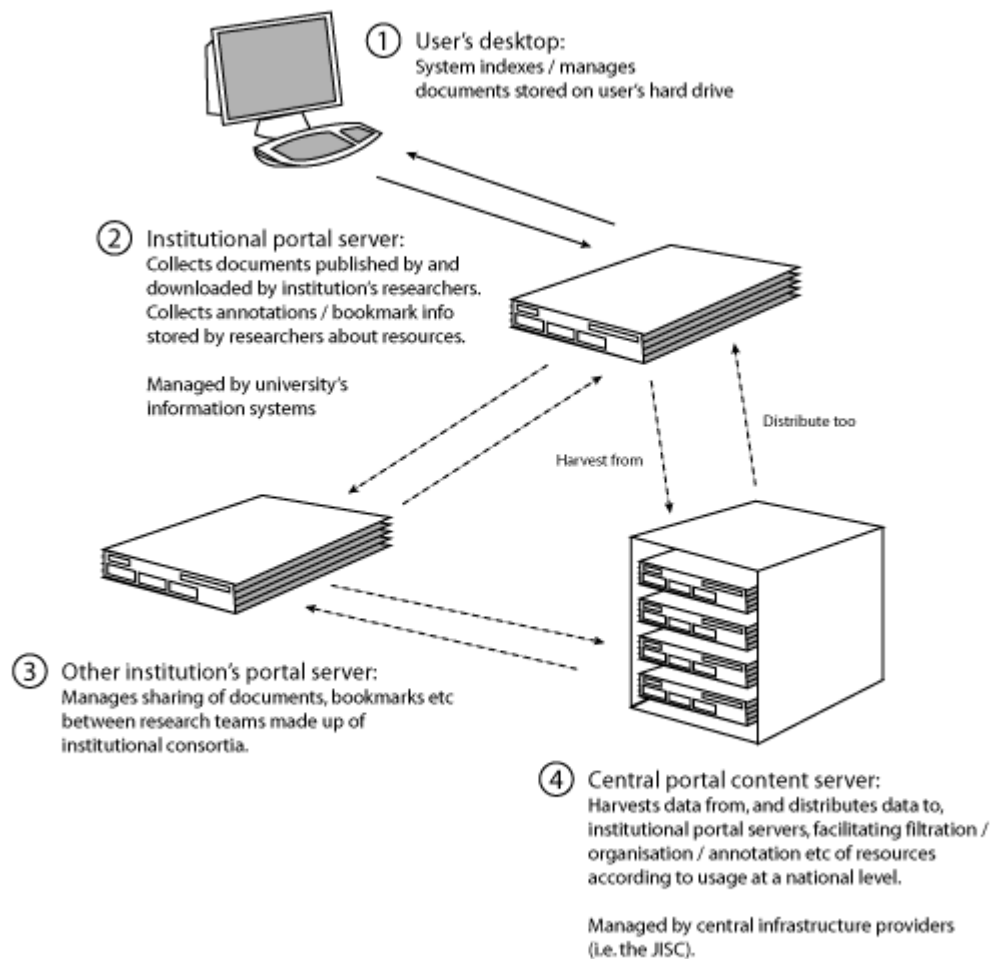

Figure 1

A8.2.1 The researcher's homepage

Figure 2

The first of the demonstrator layouts is the researcher's homepage, which (as all good homepages should) provides a general overview of the functionality of the entire system and a way into all the different areas (most of which are covered in detail in the remaining sections). The notes below refer to the key in the diagram above:

1. The **search bar** at the very top of the page would probably be a "tool bar" of some sort in the researcher's browser rather than part of a web page. It would work in exactly the same fashion as the search bar in the Firefox browser (from which the idea originates), with a field for the search term and a dropdown that allows the user to choose which of a list of search engines the term should be searched for within. Please note that the local university's library OPAC might be one of the suggested search locations. Please also note that the institution might make new options available, and the researcher might also be able to customise this list with their favourite search engines (also as per Firefox).

2. The blank space below the tool bar is a kind of "**activity bar**" which changes to include functionality relevant to the different parts of the system when the researcher is within them (covered in the remainder of this Appendix).

3. **Your documents:** this is inspired heavily by the Google desktop application, which indexes and searches documents on your local hard drive. Please note: it would probably be more useful to restrict such an index and related search functionality to a dedicated "research area" of the local file structure, so that the system isn't clogged up with shopping lists etc. In fact, it is debateable whether or not such a "research area" would be best placed on the local hard drive: it would probably be better to store them on a local, always on document server to allow easier sharing of project documents

4. **News feeds:** both Google desktop and Firefox allow the inclusion of content from standard RSS news feeds, so this system shows the same type of function (in this case showing a potential "DRHA 2006" conference news feed). The researcher would merely need to know the address of the RSS feed to set this functionality up (it's not called "Really Simple Syndication" for nothing…) This functionality would obviously rely upon the increasing provision of RSS news feeds by people such as conference organisers, the JISC, the AHDS etc.

5. Links to **bookmarks**, the researcher's **CV**, **projects** and **research partners** will be covered in more detail in the section of this Appendix that follow.

6. **Frequently used resources:** would allow researchers to browse as well as search for resources, by making the resources that the rest of the research community have been accessing available. As mentioned in Section 2.1 above, if the local research portal server were in communication with a central JISC research server (that amalgamated research resource usage data from all HE Institutions), the researcher could widen the scope of this function to see what websites, papers, datasets, books and so on researchers from across the UK were accessing. They would also be able to narrow it down to see what resources their project partners had accessed recently. The inclusion of CV and HR information would also allow filtration by academic level (e.g. postdocs) and by subject area.

A8.2.2 Search results page

Figure 3

This page design shows the first few results of a search using Google Scholar, undertaken with the **search bar** described in the homepage section above.

1. Using the functionality that has become available in the **activity bar**, the researcher has chosen to increase the number of results from the Google default of ten to 100, and has returned this set to the local institutional portal server.
2. Once the result set is held locally, it can be cross-referenced **against resource usage data** indicating how it has been used by other researchers. In this example, the researcher has chosen to order the 100 results to show those resources most

commonly visited by postdoctoral history researchers from the local institution, the University of Sheffield. Each result in the list is augmented by information about the number of researchers that have visited it, the number of references that exist to that resource in other known resources, and the number of (publicly available) notes that have been left by researchers who have accessed the resource.

This idea is an attempt to fulfil the user requirement for "quality control" of Internet resources, though it does fall some way short of the original requirement of: "… a list of sites deemed worthy of consideration after the application of a standardised set of criteria, which could be cross-referenced with the list of sites that were not deemed worthy of consideration, with the reasons why they were not included in the original list."

Given that Google searches quite regularly return result sets in the tens of millions, this request was deemed somewhat impractical! Also, there is a distinct possibility (highlighted by usability research conducted during the Aria project) that a lot of researchers might not agree with the criteria used to select "worthy" resources and would thus not trust a system of that sort, even if it were possible to set one up.

While on the subject of the large result sets that Google returns: it must be noted that this system would only allow filtration and ordering of result sets divided into blocks of 100 – 200 maximum. A system that allowed functionality such as the ability to "Search Google to show *all* results ordered by those most visited by post-doctoral history researchers at DMU" would require the retrieval and cross referencing of entire Google result sets: often in excess of 25 million results. This might be achievable, but probably only if the JISC set up this type of system *with* Google Inc. itself and licensed its entire index. To do so would most likely preclude the use of any other search engine in the system.

A8.2.3 Web page annotation
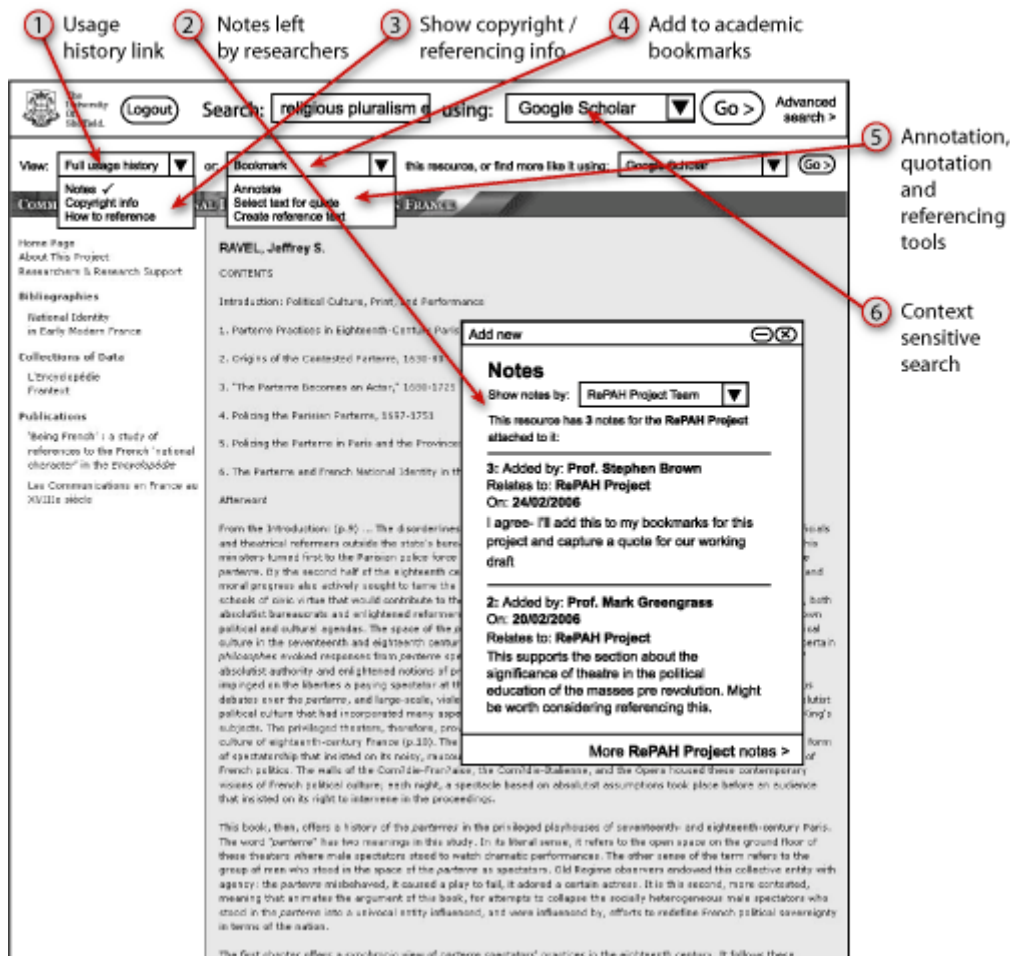
Figure 4

This design shows a web page visited by a researcher who is logged into the system. It is heavily inspired by the Diigo social bookmarking and website annotation system, for which animated demonstrators are available (as Flash movies) at:

http://www.diigo.com/help/flash_tutorial

This functionality would work by intercepting the HTML from this page at the local institution's portal server, searching for any metadata related to the URL of this page, then adding it to the original resource's HTML. This would allow for the "notes stuck to the front of the page" effect shown in the design. The demonstrator indicates the following functionality.

1. A link to the full **usage history** for this web page (see A8.2.4).
2. Any **notes** attached to this page (shown "switched on" in the design). As with other types of information, the researcher would be able to change the "scope" of

the notes to show ones kept completely private, ones shared between project partners (shown in the design), ones to be kept within the local institution (which might perhaps be the best default setting when new notes are added) and ones that are "public" to research within UKHE.

3. Access to information regarding **copyright / ownership** of the information within the page (where provided by the page's creator / publisher), alongside information about **how to reference** the page.

4. A **bookmark** function, which would add the page to their research bookmarks (see A8.3.5), rather than their standard browser bookmarks.

5. A means of **annotating** the page themselves, selecting a block of HTML text from the page to use as a **potential quote**, or generating the text necessary to **reference** the page.

6. The researcher would be able to leverage key words stored about the page (either from the page itself, or perhaps from Google etc) to perform a **context sensitive search** for more pages / resources like this one using one of their chosen search engines.

A8.2.4 Resource usage information

Figure 5

This page shows more in-depth information about a resource. It could be accessed by selecting "view full usage history" from the activity bar when viewing the resource after searching for it online (see A8.2.3), or by looking at a bookmark (see A8.2.5), or by browsing through the "frequently used resources" selected by other researchers from the local institution or UK HE as a whole (see A8.2.1). It would also be possible for the

researcher to view the usage histories of their own publications via their CV page (see A8.2.6).

Resources could be papers / articles, web pages, news stories, library books (with data held on the local OPAC), datasets etc. The page contains:

1. **Basic information about the resource** (e.g.: author, brief abstract, publication dates etc).
2. **Personal usage information** that would track when the user had first found the resource, which documents the researcher had created that referred to the resource, and which of the researcher's publications contained an official reference to it.
3. Links to **other resources that refer to this one**.
4. **Ownership** and **referencing** information.
5. **Institutional or wider UKHE usage,** indicating of how many / which other researchers had accessed the resource. The researcher would be able to change the scope of this survey (and potentially filter by subject, academic level etc) as before.
6. Functionality to compare the user's desktop version with a copy held in an institutional (e.g. University-wide) **document store**, and the "original" online version.

Notes related to a resource that previous researchers may have left could also be included in this page (but are not shown in the demonstrator).

A8.2.5 Bookmark management system

Figure 6

This demonstrator page shows:

1. A set of the researcher's bookmarks, with the first in the list "expanded" to show a fuller set of information. Clicking on the button to the right of each bookmark (viewable without expansion) would take the user to the resource itself (see A8.2.3 above), while clicking the "View full resource usage history" link at the bottom right of an expanded bookmark would take the researcher to the resource's usage history page (see A8.2.4).
2. Expanded bookmarks also show the latest note added about a resource.
3. Expanded bookmarks also link to documents and publications in which the researcher had referenced the resource.
4. Bookmarks are filterable by type of resource.

Bookmarks would not necessarily be added to this list by the researcher alone: they could also be added by research partners in reference to projects etc (not shown).

A8.2.6 Researcher's online CV

Figure 7

The idea of an online CV is inspired both by existing online academic CVs (the head of De Montfort's Centre for Computational Intelligence has a comprehensive one at: http://www.cci.dmu.ac.uk/index.php?i=5&id=1   for example) and by professional social networking sites such as LinkedIn (http://www.linkedin.com) and Ecademy (http://www.ecademy.com).

It is anticipated that this information would need some kind of input from an institution's HR Department to at least confirm details of the researcher's job title, qualifications,

career history etc (shown a 1 above). It would also be necessary to link this page to an institutional "publication repository" of the type being considered to aid the computerisation of the RAE, in order to confirm details of a researcher's publication history. Clicking on one of the publications (shown at 2) above would show the full usage history as described in A8.2.4. The page design also contains links to professional / academic associations and societies of which the researcher is a member (shown at 3 above).

This page design shows the state of the screen as the researcher administers their own CV, which means that:

4. Buttons to edit or upload new information are present, and the button to contact the researcher is greyed out. (Please also note that the researcher's contact details would not be shown on this page to prevent spamming).
5. The activity bar allows the researcher to view and roll back to previously saved versions of their CV.

As mentioned previously: it is vital to the proper running of this system that the information within a researcher's CV is made available, as reviewing a CV is an important method for fellow researchers to attach provenance to notes left by the researcher (aside from the fact this is very useful information to have available when preparing bids, attending conferences etc).

This method of providing provenance is perhaps the strongest aspect of the arts and humanities that exists within the demonstrator: other subject areas usually have other ways of attaching provenance to research (e.g. empirical testing), but in the arts and humanities the research history, reputation and track record of the researchers themselves (i.e.: how informed their point of view is) becomes much more of an issue.

A8.2.7 Project information page

Figure 8

This demonstrator page shows all the information about a particular project (in this case the RePAH project itself). Note that this page is actually a "stage down" from the "index of projects" (not included in the demonstrator layouts), which would probably look something like the bookmarks and partnership management pages.

The project information page shows:

1. The members of the **project team**: clicking their names would display their CV (as per A8.2.6).
2. Project **documents** created by all the team members and uploaded to the RePAH project folder on the local institutional portal server to allow sharing between team members. Following links to documents would display their usage history page (see A8.2.4).
3. **Email** sent between team members regarding the project. Note that the researcher can see messages they have sent, they have received and those sent to the whole team group.
4. Project **bookmarks** collected by the whole team (see A8.2.5).
5. The **activity bar** for this page, which contains a short cut to creating email, a link to the official "public" project website and a button that archives completed projects (probably only available to the project director, who would probably also be able to re-activate projects too).
6. Also shown are two **news feeds**. The first is related specifically to the project itself, and new RSS feeds could be added by all members of the project team. The second is a "general project funding" news feed, which would probably also be included in the "index of projects" page (mentioned above but not provided within these layouts), and visible on every other project page too.

The project management section of the demonstrator is very similar to document sharing and collaboration tools such as Microsoft Sharepoint:
http://www.microsoft.com/uk/office/sharepoint/prodinfo/default.mspx

A8.2.8 Research partner page

Figure 9

The final layout shows the researcher's project partners, which they would use in a similar fashion to their bookmarks (see A8.2.5). This page would:

1. View all the documents shared with their partners (see A8.2.4).
2. Link to the projects they were working with their partners upon (see A8.2.7).
3. Read email sent to and received from partners.
4. Allow the researcher to contact and view the CV (see A8.2.6) of all the partners they were currently working with.
5. View all bookmarked resources recommended by partners.

A8.3 The technologies

As noted in the above descriptions, several of the proposed demonstrator features were inspired by tools and services already available on the web. Whatever forms of ICT support structures are developed in future to facilitate arts and humanities research they will necessarily both reflect and be constrained by what is happening elsewhere on the web.

Four major developments are currently discernible:
- Internet 2
- Grid computing
- Semantic web
- Web 2.0

Internet2
Internet 2 is a US initiative to develop and deploy advanced network applications and technologies for research and higher education. Internet2 efforts are focused on:

Advanced network applications are enabling collaboration among people and provide interactive access to information and resources in ways not possible on today's commercial Internet. Interactive distance learning, remote access to unique scientific instruments, real-time access to large databases, and streaming high-definition video are all possible with high-performance networks.

New network capabilities such as Quality of Service, multicasting, and IPv6 are being aggressively tested and deployed in the networks used by Internet2 members. These capabilities support advanced network applications today, and will enable tomorrow's commercial Internet to provide the reliable performance advanced applications require.

Middleware, the behind-the-scenes software, is providing security, directories and other services required by advanced network applications. In today's Internet, applications usually have to provide these services themselves, which leads to competing and incompatible standards. By promoting standardization and interoperability, middleware will make advanced network applications much easier to use.

High-performance networks are linking the campuses and laboratories of over 200 Internet member institutions. The high-performance networks participating in the Internet2 project provide the environment in which new network applications and capabilities can be deployed and tested. For further information see www.internet2.org

Grid computing
The Grid is an architecture proposed to provide an infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources.' In this context the term "resource" includes computational systems and data storage and specialised experimental facilities as well as the kinds of data and objects more commonly recognised as resources by arts and humanities

researchers. The purpose of the Grid is to support collaborative research enterprises that require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists. Grid developments thus underpin the future of the UK eScience programme which, since 2006, includes the Arts and H. The UK e-Science Programme is fostering the development of IT and grid technologies to enable new ways of doing faster, better or different research, with the aim of establishing a sustainable, national e-infrastructure for research and innovation. The UK e-Science Programme is a coordinated initiative involving all the Research Councils and the Department of Trade and Industry with funding of £230 million from 2001-06. It has also leveraged industrial investment of £30 million. The Engineering and Physical Sciences Research Council manages the e-Science Core Programme, which is developing generic technologies, on behalf of all the Research Councils. For further details see http://www.rcuk.ac.uk/escience/.

The Semantic Web
The idea of the Semantic Web was developed by Tim Berners-Lee, web visionary and Head of the World Wide Web Consortium (W3C).

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."
Tim Berners-Lee, inventor of the World Wide Web (Berners-Lee et al 2001*).*
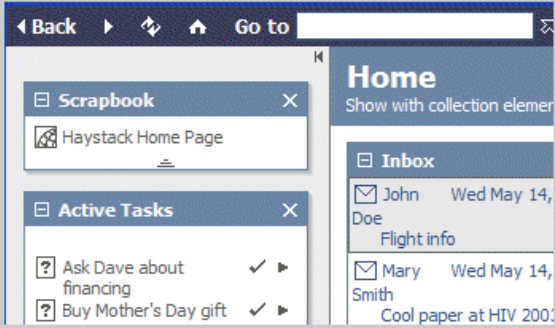
Put simply: where the "old" Internet of simple HTML could only provide links with the simplicity of "A links to B", semantic web technologies allow the crucial extra step of "A links to B *because*...They allow resources to be linked together *meaningfully* (hence "semantic web").

This really does make a lot of difference if your application relies upon amalgamating a variety of disparate online resources (or in other words: if you're making a portal). You can be a lot more intelligent in how you organise them, you can give the user more filtration and re-ordering options, for example, or you can easily flag up who put the resources there, and why. Figure 10 below is an example of a Semantic Web based Personal Information Manager, developed by MIT, that shows how this kind of data integration and filtering could work.
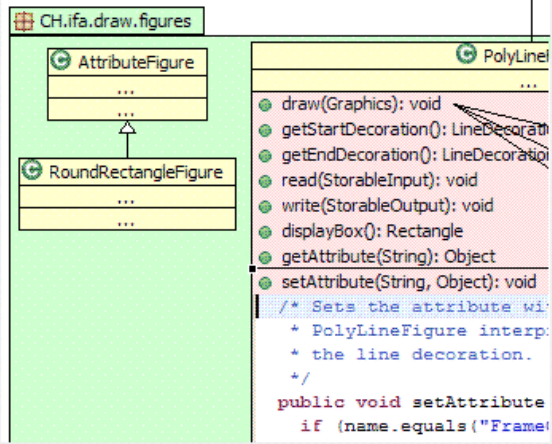
The Semantic Web relies on users being able to agree quite precise definitions of the meanings of terms or concepts and also the relationships of different terms with each other. These formal concept/relationship definitions or schema are known as ontologies and these ontologies are used to "mark up" objects or resources with metadata that web agents can then find and correlate with other resources and users' needs and profiles. Semantic Web technologies are likely to underpin the future of the next generation Web and any portal services in the foreseeable future are likely to employ Semantic Web Technologies. However, a lot of knowledge, even scientific knowledge, cannot be described in a logical way and in the Arts and Humanities, where a lot of "knowledge" is the result of heuristics and associative thinking the semantic web poses substantial

problems for use as a refined research tool. For a relatively jargon-free introduction to the Semantic Web see http://www.archimuse.com/mw2006/papers/lowndes/lowndes.html

Figure 10. MIT Haystack Semantic Personal Information Manager Source: http://haystack.lcs.mit.edu/ [Accessed 28 August 2006]

Web 2.0

Web 2.0 is a new phenomenon identified by Tim O'Reilly in the wake of the dotcom bust. He defines Web 2.0 as

- Services, not packaged software, with cost-effective scalability
- Control over unique, hard-to-recreate data sources that get richer as more people use them
- Trusting users as co-developers
- Harnessing collective intelligence
- Leveraging the long tail through customer self-service

- Software above the level of a single device
- Lightweight user interfaces, development models, AND business models

Some researchers are already familiar with popular Web 2.0 services such as Flickr (http://www.flickr.com/), Delicious (http://del.icio.us/) and MySpace (http://www.myspace.com/). Flickr allows users to upload images with descriptions, annotations and associated metadata ("tags") via an easy to use web interface. Image collections can be kept private or shared with a specific group only, but Flickr is largely a social space where people share images and comments and the more users, images, comments and tags there are, the richer the resource becomes. The idea of "social tagging" aka "folksonomies" is currently being explored by the steve.museum project. "Steve" is a collaborative research project exploring the potential for user-generated descriptions of the subjects of works of art to improve access to museum collections and encourage engagement with cultural content. (http://www.steve.museum/)

Delicious is a forum for storing and sharing web bookmarks. Again users may keep their bookmarks private or share them with selected individuals only but the main reason for the popularity of Delicious is the ability to share, link to and browse other people's bookmarks to create vibrant overlapping communities to shared interests.

MySpace is unashamedly a social phenomenon in which individuals create home pages with personal profiles, interests, diaries, comments from visitors, links to their MySpace friends, to other websites, to music, etc. It acts as a virtual space within which to meet people based on published personal information e.g.("I like eating hot chillies"), tracked behavioural patterns (e.g. number of times different discussion for a were accessed or music files were downloaded), and emergent characteristics (e.g. number and type of other people who create a link to your page from theirs).

A brief explanatory note is in order here to dispel any potential confusion between the terms "Web 2.0" and Web services" Web 2.0 is a conceptual model of how the Web can be, while Web services are a set of technical standards for creating/combining services delivered by the web. A simple example of a Web service would be an online travel agent that combines individual applications such as a calendar, a shopping trolley, an image bank, a currency converter, etc. into a single, coherent service that serves a particular purpose. The confusion arises because sites such as Flickr are providing services and they are on the Web. Even Tim O'Reilly, who coined the term Web 2.0, explicitly links it to Web services: "People don't often think of it [Web 2.0] as "web services", but in fact, ad serving was the first widely deployed web service, and the first widely deployed "mashup" ." Source: "what is Web 2.0?" [http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1] . The point to remember here is that a Web 2.0 service may not necessarily have been built using Web services technology (although it probably was) and a Web service (e.g. the online travel agent) is not necessarily an example of Web 2.0.

None of the above developments are happening in isolation from each other. We shall need the increased bandwidth of Internet2 and the shared processing and communication

tools of the Grid to fully exploit some of the potential of future Semantic Web and Web 2.0 services.

Appendix A9 Work-Package 6: Phase II User Trials of Portal Demonstrator Focus Groups and Interview Results

WP 6 Report prepared by Jared Bryson

SUMMARY

o Focus Group respondents desired simple tools that required little or no input of time or personal information. Any tools introduced must not duplicate existing systems.

o **Workflow Management tools** that give the researcher greater personal control over digital project resources, especially more evolved **bookmarking features** were identified as the most valuable. While these tools are currently available in the form of GOOGLE desktop tools, the majority of researchers were unaware of their existence, despite the ubiquitous use of GOOGLE as a web search engine. Some form of automated **copyright management system** to facilitate the growing concern with usage permission and intellectual property rights was also highly valued.

o **Resource Discovery tools** that provided greater control over web-based resources were highly valued by researchers. The ability to **filte**r the quality of hit returns, **search** multiple databases was at the top of all responses. Journal articles and online bibliographical resources are consistently seen as the most important and regularly consulted online resource by most arts and humanities researchers. The option to have comprehensive access to these was consistently the top request of capabilities that were proposed. A **web-based news feed** feature appealed to most respondents. Respondents liked the idea of a Really Simple Syndication (RSS) style system which by-passed personal email accounts, but notified users of conferences, funding, jobs and new research publications.

o **Communication tools** were not valued highly. This reflects the individualistic culture of much Arts and Humanities research. There is apparent satisfaction with existing communication systems, particularly email. Real-time 'chat' and desktop video-conferencing ranked the lowest of all tools proposed. However, **collaborative research tools** for social bookmarking, annotations and shared document editing ranked towards the middle of most responses. This is particularly interesting since several of the focus groups highlighted the lack of collaborative culture among their own disciplines.

o **Automatic information-harvesting tools** were regarded as problematic. Two automatic-harvesting tools were suggested: a) an automated monitoring of electronic resource usage by research practitioners (to assist in shaping user-needs for the future), and b) an automated harvesting of CV details to provide the basis for a national register of research practitioners. There were issues concerning the infringement of personal privacy, the challenge to a predominantly individualistic scholarly culture, and a worry among early-career academics about its possible abuse for promotion purposes that overcame the potential benefits of such automated-harvesting tools.

A9.1 AIM

Building upon the first phase of the research project, this second phase of focus groups and interviews was designed to test the value of various web-based capabilities. We used the visual props of web-style screen shots and also presented verbal scenarios of the uses of various tools. The idea was to prompt Arts and Humanities researchers to assign value

to advanced portal tools, even for those that did not, in reality, currently exist. The screenshots were therefore not functioning specimens. They did not afford the actual ability to work with these tools in real time, to manipulate data, or to work in a collaborative environment. The exercise was a 'thought-experiment', and, faced with a functioning reality, our respondents may have behaved rather differently. The purpose here was to determine those features that would be most valuable in a virtual research environment for the arts and humanities research community. The following report presents the choices made by our second set of focus groups to some possible tools development, supposing it were to become available.

The Advanced Portal features that we chose to investigate further were based on the results from the first phase of the project. They are tools that might assist a researcher based in a UK HEI to perform their research-related tasks more effectively. The challenge for the Arts and Humanities research practitioner is (as Anderson, et al noted at the All Hands Meeting in 2005):

> *not […] a data deluge in the sense used within the sciences, but rather it is the existence of a multitude of data, widely distributed, created and made available using different technical and metadata standards.*

As noted in the **Demonstrator Description Report (A9)**, the tools proposed held two prerequisites for users:
o That their use of the Internet (and other resources) for research would be monitored and indexed.
o That their research output and career level / standing would be registered and documented in a standardised fashion.

As will be seen below, these prerequisites were highly problematic for respondents, and coloured many of the responses to tools which depended on their use.

The following table presents the results of eleven portal features or capabilities drawn from the Questionnaire and Delphi exercise. The order rates the first the most valuable and the eleventh least. These features also break out into three broad categories:
o work-flow management
o resource-discovery and interoperability
o communication and collaboration.

Combined Results in aggregated order of Preference for Eleven Web-Portal Features

Resource Discovery Tools	
1.	Access to all journals
2.	Cross-database searching
3.	Pushed alerts
4.	Quality Control, Ranking and Filtering
5.	Aggregation of data
Workflow Management Tools	
6.	Personalisation and book marking
7.	Peer review
8.	Copyright management
Communication Tools	
9.	Online collaboration tools
10.	Grid Connections
11.	Desktop Video Conferencing

Figure 1

These eleven features do not map exactly onto the thirty-six features presented in the eight screen shots. The next table lists the top ten portal features from the web-page demonstrators. The discontinuity between the two sets of features is noteworthy. In this table, apart from the visible annotations as a collaboration tool, and the filtering as an aid to resource discovery, all the others would be categorised as workflow management tools. As will be discussed further below, a number of the tools appeared in slightly different forms on multiple pages. As a result the concept of an automated copyright management system ranked twice in the top ten.

Combined Preferences from Focus Groups

Top 10 Combined Tallies for Portal Features From 8 Web-Page Screen Shots	
Rank	**Feature**
1.	Keyword search of personal bookmarks
2.	All resources bookmarked
3.	Visible annotations
4.	Referencing system
5.	Frequently Used Resources
6.	Context sensitive searching for similar pages
7.	Desktop indexing & searching
8.	Filtering
9.	Copyright permissions
10.	Copyright details/ borrowing permissions

Figure 2

A9.2 METHODOLOGY

Since the focus groups and interviews were trialling 'mock-up' tools, the focus groups were given an explanatory visual presentation (MS PowerPoint) that was keyed to a paper-based evaluation form. The evaluation forms supplied to the focus groups and interviewees consisted of two exercises (**Appendix 11**).

1) The first asked the respondents to list in order of priority eleven capabilities that a digital tool might be able to provide. These were each illustrated with five hypothetical scenarios.

2) The second consisted of a series of eight wire-frame screen shots that incorporated as many as eight different web-pages. Some of these capabilities appeared on more than one web page, though they were meant to be used in different environments. For instance the presence of Really Simple Syndication (RSS) news feeds on both the researcher's home page where it might be used to receive job alerts and in the shared information of the project webpage where it might notify project teams about funding or conferences. Respondents were asked to identify the various tools' values on a five-point Lichert-scale, with five being the most valuable. Free text space accompanied each five-point scale that allowed for additional comments.

The anonymity of all participants was assured and the sessions were digitally recorded for transcription.

Timeframe. The focus groups were based around subject conferences from early April until the middle of July 2006. Interviews were held from May until July and respondents were given gift vouchers as incentives.

Demographics. This second phase of focus groups was intended to represent all eight of the AHRC's subject panels. Sessions were therefore organised at representative annual subject-specialist conferences. This strategy had strengths and weaknesses. While it was easy to locate researchers from a diverse array of UK institutions of higher education who might fit within the boundaries of the AHRC subject panels, fitting a technology-intensive focus group within the time-scale and mood of a conference was more problematic. Two sessions were cancelled, one the result of an unforeseen interruption (a fire alarm and evacuation). Participants were self-selecting. A target of six participants was achieved for five of the eight sessions, with one consisting of only four and two others with five. There was a mix of post graduate students and early and mid-career lecturers among the population.

Focus Group Populations
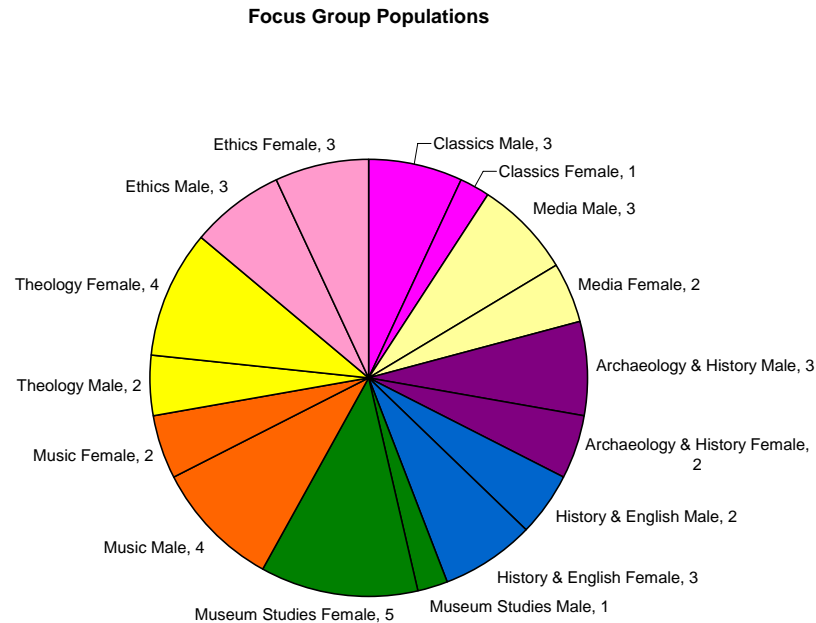
Figure 3

Interviews. Telephone Interviews were conducted with eleven individuals to supplement the findings from the focus groups and to fill subject gaps. They included interviews from these AHRC Subject panels:

- o Panel One, an archaeologists
- o Panel Two an architectural historian and an art historian
- o Panel Three a researcher in English literature and a corpus linguist
- o Panel Four a lecturer in Modern History
- o Panel Five a lecturer in French and a lecturer in Spanish
- o Panel Six a lecturer in Information studies
- o Panel Seven a lecturer in Dance Studies
- o Panel Eight two lecturers in Law

Attempts to interview representatives from the discipline of Fine Arts were, however, unsuccessful. All interviewees had the same screenshots and evaluation forms as those given out during the conference sessions. The one-to-one nature of the interview precluded, however, their being affected by the group-dynamics of the focus groups.

A9.3 RESPONSES

What follows is a cumulative description of the choices made on evaluation forms from these eight focus groups and eleven interviews. They combine both the first evaluation which ranks eleven features and the second evaluation which ranks thirty-six features, presented in eight screen shots. Where relevant, quotations which supplement the evaluation form results are included.

A9.3.1 Resource Discovery Tools.

One of the primary tasks for researchers is **locating** and **collecting** electronic information. Although the Web has been an enormous asset, Arts and Humanities researchers have consistently reported that the standard search engines are blunt instruments for searching and retrieving relevant information. Retrieval-ranking is opaquely determined. The quality and authority of the retrieved resources is problematic. Most of the important scholarly resources for Arts and Humanities scholarship are not searchable by means of standard internet search engines. So it is not surprising that the combined totals from the focus-group respondents ranked **greater searchable access** to electronic research materials as their most highly-valued feature. So **Search Control, Ranking and Filtering** featured high in the aggregate rankings. Among the respondents, Google was the search engine of choice for accessing the Web. However, the volume of worthless data returned to a general search string was often considered most problematic, if not overwhelming.

> *For example if you're doing a Google search, even if you're trying to search for something fairly specific, you're going to get a load of rubbish. And granted that they try to rank things in order of relevance and not be a lot of repetition. It's extremely time consuming and you really want to know it's searched from reliable sources rather than some wacky pressure group or something or someone's high school paper. So you want to know the results you got are worth looking at even if it turns out they're not all that you want in the end.* PHILOSOPHY 12:25

The 'authority' of what was found was uncertain.

> *I'm a bit distrustful of the Web as an information source because you don't know who the authority is.* MEDIA 18:47

Respondents wanted to filter the returns by selecting their own search algorithm, and have the ability to search for multiple elements, or at least have even greater flexibility to use search strings with Boolean limiters than is currently afforded.

> *That's quite key. The way Google ranks with interlinking and so on means that you actually get the older references. So the algorithms for ranking are quite key to how usable a tool it is. What would really be great is if you could choose the ranking algorithms yourself. So every time you do a Google search you say, rank by, newest site, not oldest site…rank the number of hits rather than the number of links.* ARCHAEOLOGY & HISTORY 29:41

Respondents did not want a search system that limited their control over or made choices for them. Many wanted to be allowed make their own choice as to what would or would not be worthwhile.

> *…any system of grading is going to be crude compared to my knowledge and long established academic ability to judge journals or judge work. The idea that it could be computerised or whatever seems improbable to me.* THEOLOGY 9:27

> *It's all very well having a ranking thing, but one of the reasons why you search for stuff is you want to find obscure bits and it's relying on somebody. I don't claim sublime wisdom about everything I'm going to find on the web but I'd rather have my own opinion and make it based on reading the article rather than having something else restrict what I actually see, and you might get a lot of dross but that's fine…* CLASSICS 25:57

> *I think this is somewhat problematic. I don't know if you had people saying this in focus groups, but to say within a discipline there are so many political stripes, so many different measures of value I wouldn't necessarily trust anyone's five-star review to tell me whether it was valuable or not for my research, and at the same time I wouldn't like my own work to be subject to that kind of scrutiny. It's already subject to the RAE and other kinds of judging mechanisms that are very complex in themselves. There's just something that makes me very uncomfortable about this ranking business.* MUSEUMS 38:57

Many respondents were worried about the potential for abuse from machine-determined ranking. They were concerned about the possibility of artificially inflating hit rates by having friends, colleagues or students visit a particular site in question.

Interoperability amongst **electronic bibliographic databases and journals** and the capability to **search across multiple databases** were the most highly-rated features highlighted by our investigation. Two issues were most clearly articulated. The first was the ability to know where reliable and up-to-date bibliographic data was to be found, including the ability to cross-search online bibliographic data in a more comprehensive fashion than that currently available through COPAC.

> *The standard database for Classics journals…there at least three or four years behind depending. So if there was something that had all the bibliographical information up to date…or fairly up to date, six months would be more useful.* CLASSICS 15:44

The second was the ability to move from a bibliographic reference to an online resource directly to that resource. **Interoperability**, in other words, was their most highly valued capability. While there was little discussion during the focus group sessions regarding this capability, the notion of searching across various databases appealed to respondents who ranked such features very highly.

Respondents did not want to see a portal system compete with existing applications. The more the systems worked together and limited the number that the researcher needed to encounter in their day-to-day activities the better.

> *I live in one house. Maybe I could live in two houses if I were very rich, but there's some similar issue. You have to…fix the holes in your roof and if these are going to be useful*

to you, you have to keep up your service. There's a limited number you can live in.
THEOLOGY 41:38

Web-based alerts. Where respondents were familiar with RSS-type individual information feeds, they valued them. It was a better alternative to an email feed, because it did not clutter the in-box of space-delimited university email accounts. Early-career and post-graduate students appreciated the possibility of receiving funding and job updates. An alerting system notifying researchers about new publication releases in their field of expertise was also mentioned as useful.

I just find it intrusive that my email is filled up and this would be so much better. It's also the conferences and job alerts that's particularly good. PHILOSOPHY 18:20

This would be easier to ignore than emails, and in that respect it would be nicer...this has got a fixed form and presumably you can tell what it is they're trying to tell you...those emails can get quite annoying, yeah? Whereas this would be quite easy.
THEOLOGY 45:35

The idea of calls for papers updates is great! (it would also be good if all journals fed into a single database and you got updates on new publications relevant to you)
HISTORY & ENGLISH EVALUATION FORMS

Automated Data Aggregators came towards the middle of the range of our users' desiderata. Their response to tools such as shared bookmarks and referencing elements was moderately favourable. They wanted interoperability, but they did not want it to be pre-determined or too mechanical.

With journals you've got things like JSTOR would it be better to look at ways of linking systems like that together on at least some sort of pan-European level.. I don't know how that works on the Continent but there are probably a lot of obscure journals that they haven't got around to digitising, but if you are just replicating another system from a standpoint that people are going to see, you would be better served by trying to create a network of databases you can access as opposed to a separate one that replicates everything. CLASSICS 17:38

A worry for me is security for my desktop/files. However, this sort of tool would be excellent for giving a sense of belonging to the group that sets it up (be that research project team, or even for undergraduates in a department), and for bringing a number of research functions together in one application (series of linked applications). LAW INTERVIEWEE

Application of Automatic Data-Harvesting Methodologies. The examples that were explored with the user groups were the possibilities of automatically harvesting forthcoming conference information, and individual CV data to create registries of current research activity.

A big advantage of this is that you could get alerts about publications. The only way that I've been able to do that is just sign up with the different publishers. You go through the thing and say these are my interests, and they will say, this book has been published, or you sign up for tables of contents. Well you have to remember to do that, but there's no central place where that's made available to the entire philosophical community and

> *that's a problem. It would be much nicer not to go to every publisher and sign up for alerts.* PHILOSOPHY 19:42

The reactions to conference information were more positive than to the harvesting of CV data. The latter feature was almost always listed among the least-favoured feature. It provoked a good deal of discussion in the focus groups. Many wondered how a synergy of participants could be built up unless everyone was required to participate in the system. Scholars from outside the UK may not want to provide professional details and could not be required to participate.

> *I see certain problems with it. For instance you'd need to have the CVs of all scholars who ever might have published a journal article or who have cooperated with the AHRC.* THEOLOGY 9:27

> *Not all researchers have a position to put up* [as a CV]. *Does this mean they are less 'worthy'?* MUSEUM EVALUATION FORM

> *This is the classic example. It's assuming that there is some time within the project to input this information. That's your early question, are you going to buy into it, are you going to put work into it initially. There is an issue there…There's a colleague of ours who did a very similar thing for European co-production funding. He built a database with the software…six months it was lying in the gutter not breathing, because the amount of initial input, the amount of critical input never happened. Not enough people bought into it. Me and my pals could have been emailing each other…there's a risk. What happened was the funding took him so far, he couldn't—it always cost more to roll out, to get it really up and running. On the face of it, some of this stuff looks fantastic, but it's the level of buy-in isn't it?!* MEDIA 1.09:20

> *I don't see the point of the CV.* PHILOSOPHY EVALUATION FORMS

One focus-group participant suggested, by way of alternative, embedding Library of Congress-style information within academic works in order to allow searching for those specific types of resources.

> *…the cross data base searching and aggregation of data…the ability to find something efficiently and narrow it down. And I would love to see something like the Library of Congress system for labelling the contents of web pages…some sort of system for standardising what's included on web pages. You'd never get everybody to use it, but you'd get serious sites, museums, educational institutes, to follow it, if it were simple enough and efficient enough. Something that would make searching more focussed.* MUSIC 21:03

A9.3.2 Workflow Management Tools.

This category of features was intended to demonstrate the possibility of gaining greater control over the resources and materials used on a day to day basis by researchers within the arts and humanities. This primarily meant that digital objects (documents, presentations, databases, spreadsheets, audio or video file) could be located and used with greater efficiency

Comprehensive book marking, desktop indexing and searching were features that appeared in various forms in several of the web page screen shots. Depending on the context of the page, each time the respondents gave the tool a different value rating. However, the ability to store and find all digital objects easily, whether created through one's own PC or web sites was recognised as a critical part of the researcher's routine. These features held immediate appeal for focus groups and interviewees alike and were understood to make life much easier for big projects handling large volumes of data or even singular projects with long time scales. Respondents were aware through the use of web bookmarking what this feature entailed. Some were also aware that Google's Desktop feature indexed personal resources in order to enable its search facility. Therefore there was little discussion during the focus groups and interviews about this capability, but a high degree of value assigned on the evaluation forms.

Annotating tools were also positively rated. Users warmed to the possibility of attaching notes to a digital resource in a wider range of formats than available at present. This was for private research purposes as well as collaborative research practice. There were some concerns expressed, however, over the degree of visibility of the annotation accorded to the viewer. Digital annotation was more positively ranked if the capacity to make the notes private or public was clearly a choice within the user's hands. Several participants across the subject-fields suggested a commentary or rating system similar to that for Amazon.com for material within shared bibliographical resources that would permit a research-community ranking of resources gradually to assemble. The form that this might take was unclear.

> IT'S THE VALUE ADDED ANNOTATION THAT YOU FIND USEFUL?
> *Yeah I find that really useful.* THEOLOGY 31:57
>
> *It's* [annotations] *basically letting us use our own language to remind us.* THEOLOGY 58:57

Personalisation and book-marking. This family of features allows the researcher to become more directly their personal manager of online digital libraries, storing references to materials by means of bookmarking, then (eventually) being able to index or key-word search the bookmarks, and eventually the items themselves. The ability to customise and control this process was positively-rated among the future tools for development. Users particularly welcomed the possibility to annotate the bookmark links with abstracted information or additional material so that it might form an annotated link or 'note' to a resource.

> *I always have difficulty finding specific books and if you computerised it that would be fantastic. If you bookmarked it and could get that straight away and you wouldn't have to remember the search you went through to get that* MUSIC 16:57
>
> *Actually something I've always wanted to see is a sort of two-stage book mark where you have your most frequently used resources at the top but then everything else that ends up that you book mark ends up alphabetised or something like that. It seems now you can do one or the other but I can't have a section of the five that I use the most and then the next section be the forty-nine things that I don't want to loose track of that I use four times a year and some way of compartmentalising sections of those things.* MUSIC 29:26

> *Keyword search of content of personal bookmarks would be good.* INFORMATION STUDIES INTERVIEWEE

> *You have to have a personalisable interface so people can choose which one of your facilities that they want displayed…there's got to be some way that people have got control over their public face in a way, because we're all really, really busy, and we're all trying to look as professional as we can be.* INFORMATION STUDIES INTERVIEWEE 28:41

> *Some sorting might be useful here, especially for a long list of bookmarks. Bookmarks could be grouped according to format (i.e. websites, books, articles etc.).* FRENCH INTERVIEWEE

Automated Bibliographical Downloads. Most of our users were familiar with the Endnote bibliographical software, even if they had not used it themselves, or had not availed themselves of all its features. They responded positively to the broader application of selective bibliographical downloads, with investment in making the download filters simpler and easier to use regarded as a priority. Users were also positive towards the notion that automated bibliographic downloads might include references to other places where the work had been cited. **Citation history** was also regarded, in a positive light. Though available in several of the existing e-resources and even Google Scholar, embedding a similar system with all the other searched material was considered very valuable.

A 'Frequently-Used Resources' Tool. This was variously regarded. Some respondents wondered how such a feature differed from simply storing resources on the desktop.

> *What does 'Your Documents' do that Windows doesn't do already?* MUSEUM EVALUATION FORM

> *'Your documents' is a common feature in Windows, would it be needed here?* HISTORY & ENGLISH EVLAUTION FORM

> *The Frequently Used Resources, the problem with that, and again this is speaking from my own laziness, if it only gives you the top five then I'll never use anything else.* MUSIC 28:45

However, others thought a customisable ranking feature was a positive tools development. They readily appreciated the advantages of web-based resources being accessed more conveniently and organised around common tasks. By contrast, a **'Resource Use Tracking and Usage-History Tool** was much less highly valued, in fact several thought it might be a problem.

> *Resource Usage History, unless it was monitored and controlled fairly carefully it really would be terribly open to abuse.* ARCHAEOLOGY & HISTORY EVALUATION FORM

> *It is useful to see resource usage history but I do find statistics can be somewhat arbitrary and I would be concerned if they were used e.g. in the RAE as evidence of esteem.* SPANISH INTERVIEWEE

> *I don't see why any of this is necessary at all actually. This particular feature seeing that we can use other software already why should we should we be part of this Big Brother*

publication of everything that is going on? Why should other people be able to check up on us? MUSIC 43:29

Institutional and citation history information would be valuable for the interdisciplinary aspects of some of the topics I research. Different departments tend to work in isolation, although there are valuable points of overlap between our research areas (and methodological approaches). The facility to see who else accessed material could help highlight others working in relevant subjects. DANCE STUDIES INTERVIEWEE

One respondent from the Philosophy focus group linked the resource usage history tool with the use of annotations in order to create a forum for debate:

If annotations function was more like an online discussion/debate then resource usage history would be more valuable. PHILOSOPHY FROM EVALUATION FORMS

The common user experience was that they simply did not make much use of tracked actions currently. One interviewee considered it a better tool for institutional libraries to track downloads from central document stores.

Resource usage history useful for institutions but not necessarily for individual researchers. VISUAL ARTS INTERVIEWEE

A 'Peer review' Tool. The notion here was a pre-print peer-review tool with a rating system that was more commonly understood and transparent within a process that could be conducted electronically. This was a tool that was positively viewed as contributing to a more readily understood, and more broadly shared sense of peer review.

How many people have reviewed it? Who are the people that are the peers? And then you're under the assumption that the really busy important people that know a lot about this stuff will be too busy to do any peer reviewing on your online system…So I'm always a bit sceptical about that kind of stuff. INFORMATIONS STUDIES INTERVIEWEE 23:40

A 'Copyright management and permission information' Tool. This feature proposed an automated electronic means for seeking copyright and permission information, deriving copyright from the bibliographic electronic data already stored. Respondents consistently placed this in their 'top ten' wish-list. They interpreted its desirability in both research and teaching contexts.

A9.3.3 Communication Tools

This category of features polarised our users. On the one hand, the ability to share documents and annotate resources was considered highly valuable. However, real-time 'chat' and 'desktop-conferencing' scored at the bottom of the scale. The low value of these features correlated with the satisfaction expressed in the focus groups and interviews with current communication arrangements for research purposes in these domains.

'Document-sharing' Tools. This feature was consistently the most positively-weighted of the communication tools proposed. The possibility of being able jointly to edit a

document and control the versions produced attracted our users, with the caveat that they were able to control how the sharing occurred:

> *For me sharing documents is one of the most useful things here*. THEOLOGY 47:50

However, a contrasting argument was also voiced:

> *A general worry that I have has to do with making certain aspects of the research process a public event. So things like…on the resource usage information page and it shows you everybody who has accessed an article, I just really don't like that. What does it matter if I've read the article or not. I don't want my colleagues necessarily to know if I've read the article or not, because there are obvious reasons why people would want to know that information potentially in some cases, not in all cases that it's unlikely. I just think there are certain aspects of this where we're edging into that territory where everything you do is scrutinised and it's worrying. And there's a lot of it that's very, very useful, but there's a lot of it that's edging along toward questionable…I think there will be resistance to it. People are already resistant to using the JISC's systems, and maybe that's just 'old-fogeyism' and maybe that will change as…I don't know, it's changing the nature of what research is, especially the humanities*. MUSEUMS 1.09.56

Online Collaboration tools. Social bookmarking, live chat, and group working environment tools were not positively regarded among our respondents. Their negative responses were governed by questions of time-management and utility.

> *Social bookmarking—only if you can moderate who can take your bookmarks*. HISTORY AND ENGLISH EVALUATION FORMS

> *The value of social bookmarking depends on how 'generous' researchers want to be. They may only want fellow research collaborators to see their work. They may not want the risk of others 'pillaging' their work for their own ends*. ARCHAEOLOGY INTERVIEWEE

> *Although I don't do much collaborative work, particularly outside my own institution, access to resource bookmarks of colleagues would be beneficial*. DANCE STUDIES INTERVIEWEE

Many of the respondents claimed that they did not work collaboratively and that the concept was not 'normal' in their discipline.

> *The problem is there's not so much collaboration for the most of us*. THEOLOGY 47:35

> *Theology isn't famous for being a collaborative subject*. THEOLOGY 47:54

> IF YOU HAD THE TOOLS THAT ENABLED YOU TO COLLABORATE WOULD YOU COLLABORATE MORE?   *It would take time to learn a culture of collaboration if I'm honest*. THEOLOGY 48:03

> *The sciences, if you see a paper with scientists you see a paper with twenty names to the top of the paper. If you see two at the top of a humanities paper it's a sign of an unnatural relationship. It just doesn't happen*. CLASSICS 49:20

> *Collaboration is 'made up' because that's where the government funding seems to be going, but it isn't real, you know. People do their own thing as much as possible*. IF THE TOOLS WERE AVAILABLE WOULD THAT ASSIST IT IN ANY WAY? *The nature of the work isn't collaborative. It's not like science based things where you have*

> *ongoing--you have your idea and you have to find it there yourself, you don't want to give it to somebody else.* CLASSICS 44:39

> *I would be worried if the AHRC made it a condition, a compulsory condition to engage in this. Obviously they'll pilot it...You work for them for research and have to engage with partners...it seems you're forced into a condition of sharing and that there is this idealised view that people do want to share.* MUSEUMS 1.04: 17

Desktop Video Conferencing consistently ranked at the bottom of choices. **Real-time 'chat'** was already available to researchers who wanted it in applications such as Microsoft Network (MSN). Our users told us that they did not currently use it, however, for pursuing their research and teaching. **Archiving.** This feature was consistently ranked in the lower third of responses. The low rating may have been a consequence of where the feature was positioned in the screenshots. Listed among the Project Information Page tools, there was no discussion when this feature was presented, and although it may have been received better than the worst received tools on the page, chat and video-conferencing it remained lower than bookmarking and document sharing which had been identified on previous pages as being valuable.

'Institutional Repository' Tool. This was one of the 'automatic harvesting' features that we proposed, in this instance providing a tool for user-controlled ingestion of electronic material to an 'institutional repository'. Users were not very well-informed about the 'institutional repository' movement. Their responses were governed by their belief that this form of publication was simply not a priority for them.

'Grid Connectivity Tool'. Accessing the Grid was presented as an 'infrastructure-enabling' tool to our respondents. None of them had used the Grid for their research. Their awareness of the E-Science and E-Research agenda was varied. The low ranking of this feature is no doubt a reflection of the fact that the Grid is not yet regarded as an important arena for Arts and Humanities research.

A9.4 ADDITIONAL COMMENTS AND DISCIPLNIARY DISTINCTIVES

Many respondents echoed concerns from the first phase of focus groups. There was a great desire for **simplicity** such as that found in Google's single line search field. Tools should not be laden with jargon and should not require a great deal of time in training and familiarisation.

> *They want their black box. They don't want to know what's in the black box, they just want it to work.* INFORMATION STUDIES INTERVIEWEE 8:07

> *Some people don't even read their emails, so they haven't even got past that yet!* MUSEUMS 1.04:04

> *This is becoming way too complicated...I just want it to do the job I want it to do.* THEOLOGY 34:48

> *I keep thinking is this necessary? I've got so much I've got to do. Theoretically I can see that this could be good that it can help me organise and could even make my life easier,*

quicker but I just want to run screaming from this room and say, 'Oh my God, no!', because it just seems to me to be just more things on top of what I'm doing. More things I'm going to have to learn how to negotiate, which actually normally I don't mind. THEOLOGY 39:30

I am very good at technology but don't give me this jargon. I don't understand it and I don't have the energy…I just want to do my research not this. If it furthers what I have, give it to me. If it's just going to replicate, or confuse me, or take my time up then no, I'm busy. THEOLOGY 1.07:03

My basic comment is that I'm overwhelmed by this ability, possibility, on the other hand there's always a certain time limitation which prevents you, would prevent me from using something like this. MEDIA 1.00:14

The **training implications** for complex tools are an important issue.

There are certain colleagues who have tremendous energy and enthusiasm for the new web technology and then there are others like myself feel themselves sufficiently busy and are bothered by the start up cost and may not be visionary enough to see what the grid could be. So I can imagine that you'd want to work, if you could, work with that select group which will be a real minority, maybe ten percent or less who have a real enthusiasm for cutting edge technology, who then might share the success with the rest of us that there might be some good to be done. Whereas people like myself are saying, this is really complicated and I'm ok as I am. Which I can imagine in ten years time I'll be thinking how useful this is, but I'm not one of those who can get enthused about it in advance, and until I've actually seen someone operating it. THEOLOGY 15:15

It's a question of how much time do you envision people spending on this. There's a danger that you can spend time constructing an elaborate system that will actually take over. We have this at uni where we have this way of keeping track…you've got all your activities, all the things you've learned from, all the things you hope to learn from them and any issues you have, any thing and everything has to be logged, and all data that's stored and supposedly you can go back and modify it. The question is whether time is best spent in a library with a book or looking at data online say, rather than actually ploughing through this sort of thing. Alright it can make your life a lot easier in some cases but you always have to ask does it help you do research in the most fundamental sense of the word…I mean people have always managed to get their PhDs and stuff without the help of such things. And the trouble is that if it's introduced by a funding body you have the idea that this is what you have to do. In a lot of departments people get stressed just by the fact that they have to do such things, and it annoys them because they want to go and play with pots or whatever. So it has to be quite reflexive. You say you minimise the things you don't want, but it has to take up as little time as possible because it's always an ancillary to research. Collaborative stuff might happen at the post-doctorate level but you wouldn't want [to], *I've had friends whose bits of work have been nicked and published and things and people are very protective of what they do, understandably, and it's tailoring that for subject specific* [work]... CLASSICS 47:01

Many respondents did not want to see **duplication** of applications.

Is this doing anything other than what I already do? And it seems to me we're replicating rather than helping…I am getting a bit concerned about the amount of money that is potentially going in to doing this when all I'm going to do at the end of the day is

> *minimise, minimise, minimise. As someone trying to apply for funding off the AHRC, give it to me instead!* THEOLOGY 1.04.:07

Some respondents touched on questions of **data protection and personal privacy** with regard to data sharing and the machine logging that was needed to use some of the collaboration, cross-database searching, and aggregation tools.

> *Issues of research confidentiality and issues of intellectual property rights which could arise out of sharing details* HISTORY & ENGLISH EVALUATION FORMS

> *Why is this necessary?! I don't want people to see what I do without my permission—I am a person and a scholar not a web-page.* MUSIC EVALUATION FORMS

Customisation and control of the search, storage and retrieval process was also a key concern.

> *As long as I could minimise it so it's not in the way and I can edit things and take out what isn't, because I can't stand having junk on my computer screen I don't need. So as long as I could take out news feeds because I didn't want it for the next two months and then stick it on when I do. So as long as I have control—this I understand. This I like.* THEOLOGY 24:15

No one wanted to see a new system imposed on the arts and humanities research community.

> *I wonder if there's an underlying technological issue which is people who work in the AHRC think of this as our natural home, so that we would want to work in an environment that was determined by the AHRC and get into the way the AHRC looks at things, whereas actually the AHRC is a wonderful body who gives us money sometimes and who we deal with when there is there's any chance of getting it, but otherwise there are lots of other homes.* THEOLOGY 1.06:20

The disciplinary distinctiveness emerged, albeit within the common framework of responses already outlined. Those within **Classics** (excluding classical archaeologists) claimed that their discipline did not routinely collaborate, but instead rewarded the solitary scholar working within a small network of colleagues. Their other concerns centred on ease with which the technology could be used and not be a distraction from their primary work with ancient texts. There was a lack of awareness of the potential of ICT to enhance their research. The **Archaeology and History** focus group consisted of researchers already familiar with humanities computing and its application to the discipline. They placed most value on greater development of data aggregation and cross-database searching. The referencing system and document sharing features also ranked highly. As might be expected the **Media and Film** respondents were interested in the ability to incorporate video materials as a resource. This could include searching and storing video resources with the same ease as that of text-based resources. They also mentioned increasing the ability to network for both teaching and funding purposes. For the **History and English** focus group one of the interesting features discussed was an alerting system dedicated to bibliographical information. An RSS feed for new books within the field of the researcher's interest. The **Museum Studies** focus group highlighted their unique position straddling both higher education institutions and those

outside. It was noted that many museums do not even have web access and would not be able to take advantage of the proposed tools. The CV-based quality control system would not suit the career paths and research tasks of many museum professionals. Both museum professionals and archaeologists have large populations who conduct their research outside HEIs and their ability to take advantage of the portal might be limited. **Music** respondents placed an automated copyright management system at the top of their valued features. In addition, theirs was the only focus group to highly value a system that linked researcher's published material to an institutional repository. The **Ethicists** suggested that web-page features that brought together the collaborative annotations and a blog-style chat would be very valuable to create tools for debate and a forum for sharing ideas. **Theology** respondents placed the notes and annotations twice within the top ten of the total thirty-six.

A9.5 CONCLUSIONS

Overall the focus groups were positive about the potential that the proposed tools offered; however that general enthusiasm was tempered with the caveats already mentioned in the additional comments section. The overall picture of priorities that emerges is:

o Focus Group respondents desired simple tools that required little or no input of time or personal information. Any tools introduced must not duplicate existing systems.

o **Workflow Management tools** that give the researcher greater personal control over digital project resources, especially more evolved **bookmarking features** were identified as the most valuable. While these tools are currently available in existing forms such as GOOGLE desktop tools, FLIKR or Delicious, the majority of researchers were unaware of their existence, despite the ubiquitous use of GOOGLE as a web search engine. Some form of automated **copyright management system** to facilitate the growing concern with usage permission and intellectual property rights was also highly valued.

o **Resource Discovery tools** that provided greater control over web-based resources were highly valued by researchers. The ability to **filte**r the quality of hit returns, **search** multiple databases was at the top of all responses. Journal articles and online bibliographical resources are consistently seen as the most important and regularly consulted online resource by most arts and humanities researchers. The option to have comprehensive access to these was consistently the top request of capabilities that were proposed. However, respondents also consistently wanted these features on their terms, gaining greater control over the searching process and reticent to contribute personal time and information to learning a new system. The two requirements set for many of the features of contributing professional credentialing information and time learning and setting up the system (**see Demonstrator Description Report A9**) appeared to be insurmountable barriers. A **web-based news feed** feature appealed to most respondents. Respondents liked the idea of a Really Simple Syndication (RSS) style system which by-passed personal email accounts, but notified users of conferences, funding, jobs and new research publications.

o **Communication tools** were not valued highly. This reflects the individualistic culture of much Arts and Humanities research. There is apparent satisfaction with existing communication systems, particularly email. Real-time 'chat' and desktop video-conferencing ranked the lowest of all tools proposed. However, **collaborative research tools** for social bookmarking, annotations and shared document editing ranked towards the middle of most responses. This is particularly interesting since several of the focus groups highlighted the lack of collaborative culture among their own disciplines. Following from the last quote above, the reticence to contribute personal data as well as time seem to mitigate

against not only the ability to harvest data from across the research community, but also to work in strongly collaborative environments. Despite such reactions University's are already monitoring and collecting data from scholars within their institutions. The information obtained about Professor Mark Greengrass in the dummy demonstrator data was gleaned from freely available information already available on the Web. Researchers in the arts and humanities routinely participate in 'weak' collaboration by sharing citations and interacting through their informal networks to exchange ideas and comment on each other's works.

o **Automatic information-harvesting tools** were regarded as problematic. Two automatic-harvesting tools were suggested: a) an automated monitoring of electronic resource usage by research practitioners (to assist in shaping user-needs for the future), and b) an automated harvesting of CV details to provide the basis for a national register of research practitioners. There were issues concerning the infringement of personal privacy, the challenge to a predominantly individualistic scholarly culture, and a worry among early-career academics about its possible abuse for promotion purposes that overcame the potential benefits of such automated-harvesting tools.

Within this overall picture clearly there are important differences in priorities between the groups, reflecting their particular domain research concerns and practices. Thus although 'access to all journals' ranked either first or second in value for the focus groups, apart from media and film studies, the interviewees created a much murkier picture. The dance studies respondent and the corporate linguist listed this capability last. Therefore any new virtual research environment-style portal would have to be modularised in some way such that individuals could select the features they most valued to create a personalised toolset.

APPENDIX 1 (FIGURE 4) COMBINED PERCENTAGES OF FEATURES

APPENDIX 2 (FIGURE 5)

Combined Percentages in Order of Feture Preference

Keyword search of personal bookmarks
Frequently Used Resources
All resources bookmarked (not just web sites)
Visible annotations
Referencing system
Notes
Context sensitive searching for similar pages
Copyright permissions
Copyright details/ borrowing permissions
Quality control
RSS Funding alerts
Access filtering
Resource Usage History
Citation history
Version editing
Archiving
Shared documents
Shared documents
Personal usage information
Usage history
Desktop indexing & searching
Linked to an institutional publication repository
Bookmark
Filtering
Social Bookmarking
Resource bookmarks
Communication
Resource bookmarks
Collaborative bookmarking
News Feeds
Ranking
Annotations & Quote sampling
Institutional usage information
Usage History of Page
CVs
Live chat

APPENDIX 3 (FIGURE 6)

Classics

APPENDIX 4 (FIGURE 7)

Archaeology & History

APPENDIX 5 (FIGURE 8)

Media

APPENDIX 6 (FIGURE 9)

History & English

APPENDIX 7 (FIGURE 10)

Museums
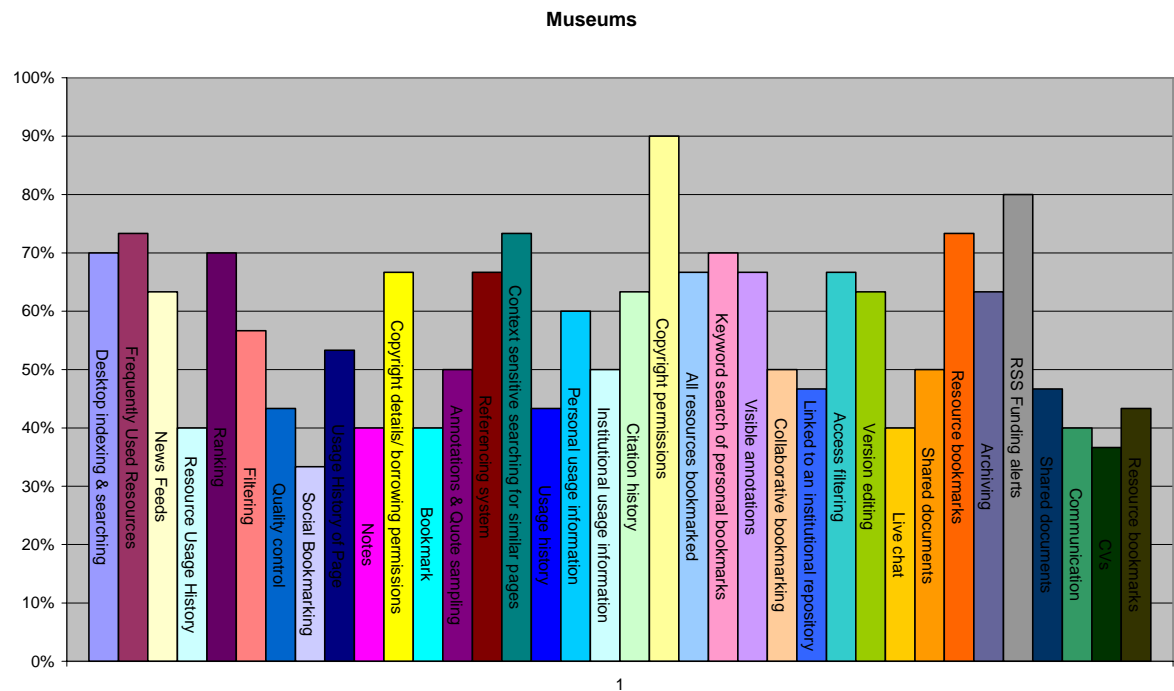
APPENDIX 8 (FIGURE 11)

Music
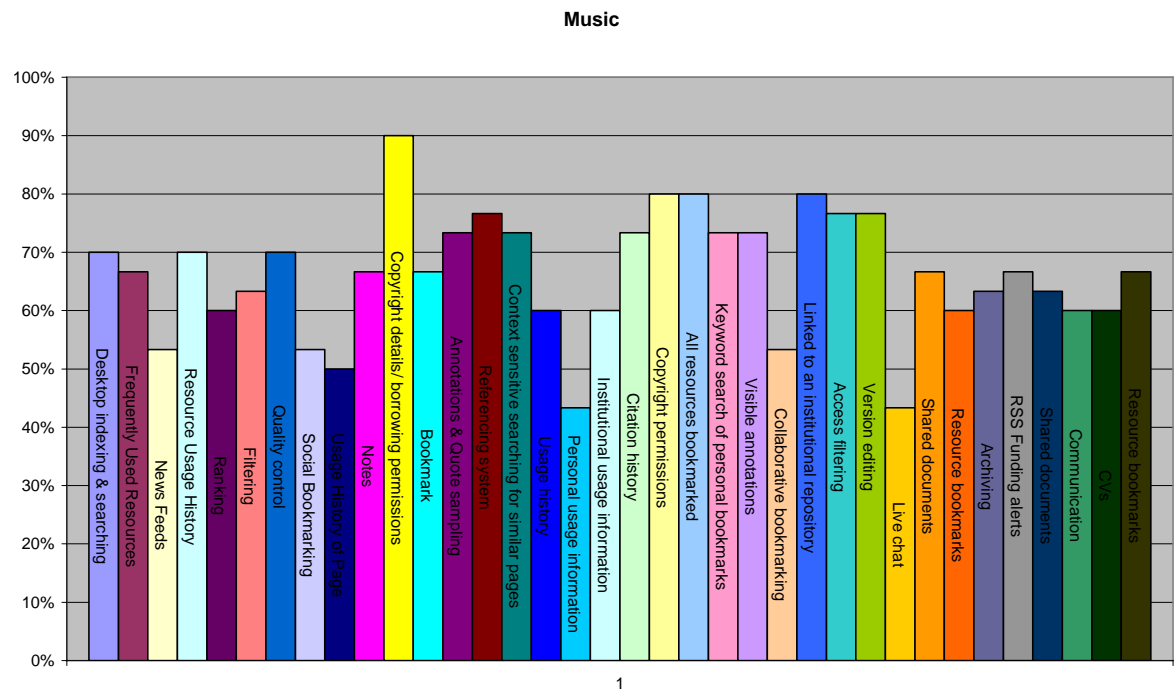
APPENDIX 9 (FIGURE 12)

Ethics
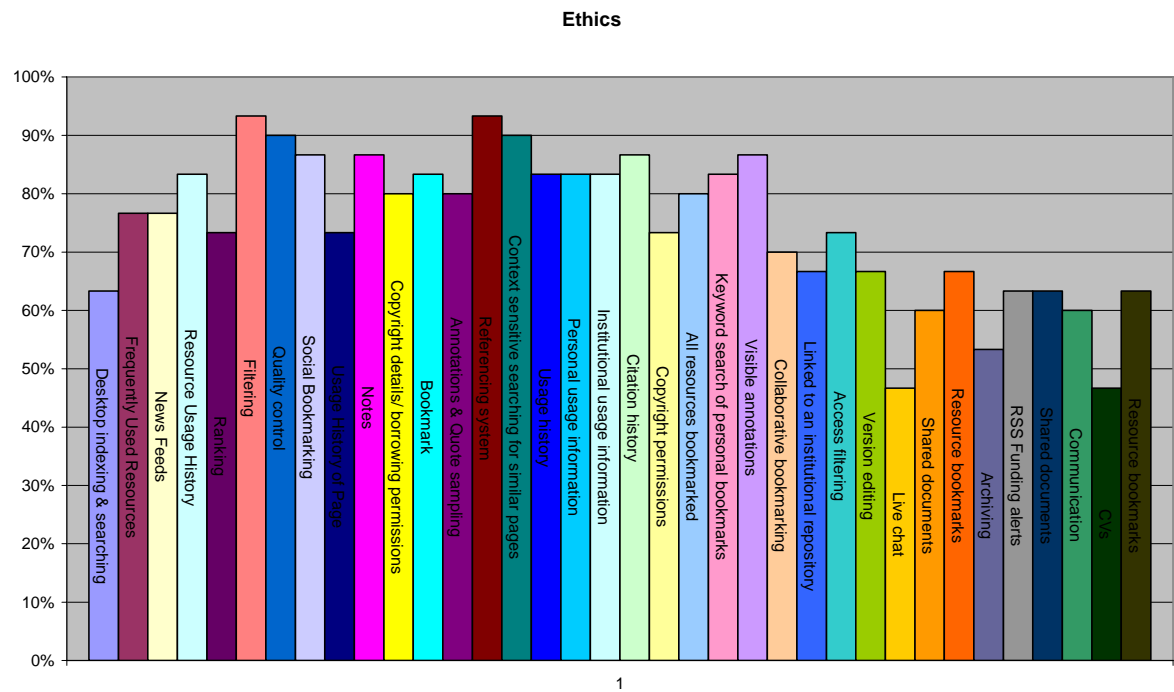
APPENDIX 10 (FIGURE 13)

Theology
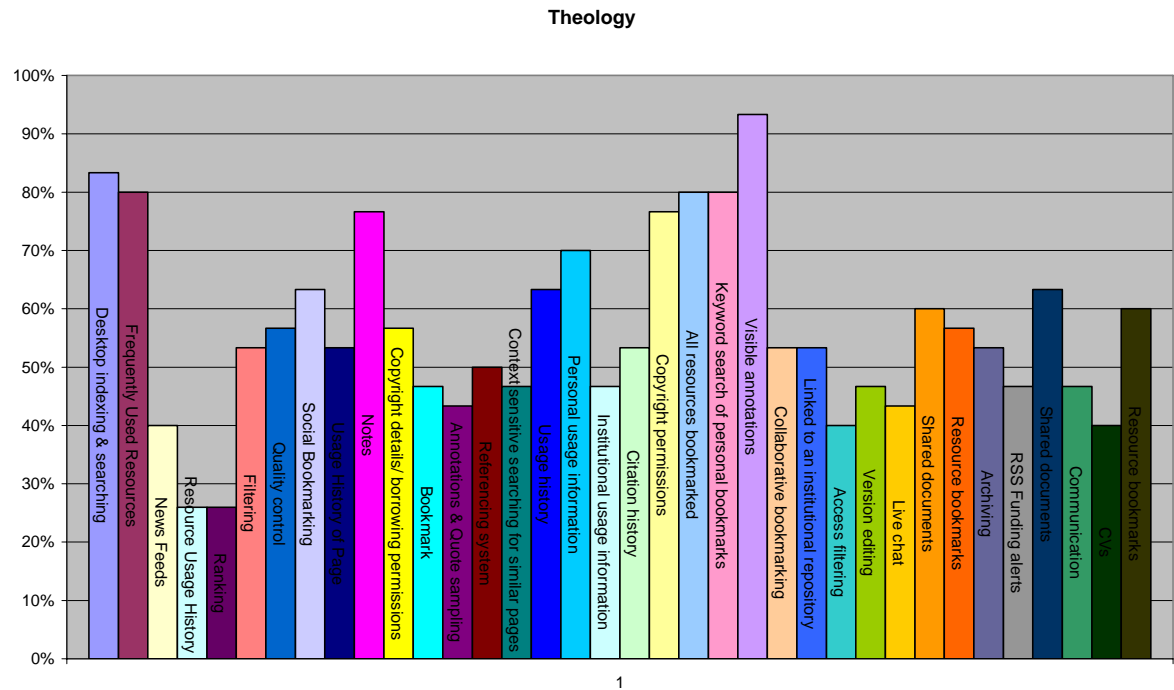
APPENDIX 11 (FIGURE 14)

Portal Demonstrator Evaluation Forms for Conference Focus Groups

Conference_____

Location_____

Date_____

Subject Panel_____

Group Numbers_____

Value rank out of 11	Portal Features		Value rank out of 11
	Cross database search	Quality control and ranking system	
	Online collaboration tools	Access to all journals	
	Desktop video conferencing	Copyright management	
	Aggregation of data	Pushed alerts for funding/conferences/papers	
	Grid connection/services	Personalisation & Bookmarking	
	Peer review facility		

Terms for Portal Features

1. <u>Cross database search</u>: *Accessing multiple databases simultaneously. These can be of different types of data and stored at multiple locations*
2. <u>Online collaboration tools</u>: *Enabling work to be done on the same set of data (or even multiple sets of data) by more than one researcher, even if they are in different locations*
3. <u>Desktop video conferencing</u>: *Using one's personal computer to conduct high-speed, high quality conversations over the WWW, rather than needing to access specialised facilities*
4. <u>Aggregation of data</u>: *Bringing different types of data, from different locations, together into one place for analysis and presentation. Data in this instance can be composed of digitised text, images, audio or video*
5. <u>Grid connection/services</u>: *This concept has several different names, such as e-science or virtual research environments (VREs), however, the overall concept is the ability to conduct multiple computational tasks very rapidly and in a collaborative environment. Computer networks are often directly linked together, thus enabling increased speed and security.*
6. <u>Peer review facility</u>: *The feature enables the data user to participate in the peer review process with anonymity and within the administrative criteria established for each particular subject specialty.*
7. <u>Quality Control and ranking system</u>: *Searches would yield web sites and journal articles with grades of reliability based on a universal standard of validation, setting the search against a list of all potential hits with reasons for not including them in the validated list*
8. <u>Access to all journals</u>: *Access to an array of primary and secondary literature, some of which may not be taken by a university library, but are nevertheless necessary and specific to a researcher's subject specialty. The portal provides access to journals including those discovered serendipitously and held by commercial, subscription services*
9. <u>Copyright management</u>: *Automatic a*dvisement concerning copyright access and use of specific audio and video downloads, offering permissions or royalty information/transactions*
10. <u>Pushed alerts for funding/conferences/papers</u>: *This feature picks up funding alerts from various sources, including research councils, government agencies, private foundations and international organisations. The same alerting service provides regular notification of conferences, calls for papers and new publications in the researcher's field of interest*
11. <u>Personalisation & Bookmarking</u>: *The ability to customise features, layout and data to suit personal needs. Easy access to large, personal bookmark library through keyword searches*

Researcher's Homepage

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Your Documents desktop indexing & searching	1	2	3	4	5
Frequently Used Resources	1	2	3	4	5
News Feeds	1	2	3	4	5
	1	2	3	4	5

Any other comments

Search Results Page

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Resource Usage History	1	2	3	4	5
Ranking	1	2	3	4	5
Filtering	1	2	3	4	5
Quality control	1	2	3	4	5

Any other comments

Web Page Annotation

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Social Bookmarking	1	2	3	4	5
Usage History of Page	1	2	3	4	5
Notes (able to be switched on/off for viewing & degree of privacy)	1	2	3	4	5
Copyright details/borrowing permissions	1	2	3	4	5
Bookmark	1	2	3	4	5
Annotations & Quote sampling	1	2	3	4	5
Referencing system	1	2	3	4	5
Context sensitive searching for similar pages	1	2	3	4	5

Any other comments

Resource Usage Information Page

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Usage history of CV	1	2	3	4	5
Personal usage information	1	2	3	4	5
Institutional usage information with filters for subject, academic level, etc	1	2	3	4	5
Citation history	1	2	3	4	5

Any other comments

Bookmark Management Page

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
All resources bookmarked (not just web sites)	1	2	3	4	5
Keyword search of personal bookmarks	1	2	3	4	5
Visible annotations	1	2	3	4	5
Collaborative bookmarking	1	2	3	4	5

Any other comments

Your CV

Logged in: **Prof. Mark Greengrass**

Your CV

Biography

Description last updated: 18/03/2003

Mark Greengrass joined the Department of Medieval and Modern History back in 1973. His research interests concentrate on the history of Europe in the post-reformation period, with particular emphasis on the political history of France and the intellectual history of Europe more generally. You may still discover copies of his France in the Age of Henri IV, (London and New York, Longman, 1995) and The European Reformation (c.1500-1618), (London and New York, Longman, 1998) on the shelves of discerning bookshops.

Update >

Current position held:

Title:
Professor of History
Institution:
University of Sheffield
Faculty:
Social Sciences
Department:
History

Edit >

Research Projects

#	Title	Started	Ended	
1:	Research Portals in the arts and...	08/2005	-	More >
2:	ARMADILLO: information mining...	08/2005	-	More >
3:	The Hartlib Papers Project.	03/2002	09/2004	More >
4:	John Foxe Project.	11/2001	-	More >
5:	Governing Passions: the reforma...	10/2001	06/2003	More >

More projects ▶

Associations:

1: The European Science Foundation

2: The Royal Historical Society

3: The Society of Antiquaries

4: The Hugenot Society for Great Britain and Ireland

Add >

Publications

#	Title	Published	
1:	Overview of the use of digital resources in t...	18/01/2006	Usage Info >
2:	Samuel Hartlib and universal reformation: s...	07/12/2005	Usage Info >
3:	Miracles and the Peregrination of the Holy i...	19/08/2005	Usage Info >
4:	Samuel Hartlib and the Commonwealth of L...	13/03/2005	Usage Info >
5:	Informal Networks in Sixteenth-Century Fre...	18/11/2004	Usage Info >
6:	Financing the Cause: protestant mobilisatio...	07/07/2004	Usage Info >
7:	Conquest and coalescence	19/05/2004	Usage Info >
8:	The Letters of Cheney Culpeper, 1641–1657	25/02/2004	Usage Info >
9:	Processing morphological variants in searc...	08/11/2003	Usage Info >
10:	France in the age of Henri IV: the struggle ...	06/10/2003	Usage Info >

Add new >   More publications ▶

Career history:

Job title:
There comes a point where it's best to stop making things up!

Held from: 1969
Held to: 1970

Add >   Position 1 of 6 MORE>

Publish all changes >

Terms and conditions of use   Help   Send feedback

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Linked to an institutional publication repository	1	2	3	4	5
Access filtering	1	2	3	4	5
Version editing	1	2	3	4	5

Any other comments

Project Information Page

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Communication live chat	1	2	3	4	5
Shared documents	1	2	3	4	5
Resource bookmarks	1	2	3	4	5
Archiving	1	2	3	4	5
RSS Funding alerts	1	2	3	4	5

Any other comments

Research Partners Page

CIRCLE NUMBER FOR VALUE OF FEATURE
(1-LOW VALUE 5-HIGHEST VALUE)

Page Features					
Shared documents	1	2	3	4	5
Communication	1	2	3	4	5
CVs	1	2	3	4	5
Resource bookmarks	1	2	3	4	5
	1	2	3	4	5

Any other comments

Appendix A10 Work-Package 7: Intute In the Light of This Report

WP7 Report prepared by Jared Bryson
Additional data supplied by Jayne Burgess, Intute-Arts and Humanities

A10.1 What does Intute-Arts & Humanities do?

> *We are a free online service providing you with access to the best Web resources for education and research, selected and evaluated by a network of subject specialists. There are over 18,000 Web resources listed here that are freely available by keyword searching and browsing.* (http://www.Intute.ac.uk/artsandhumanities/)

Intute operates on the premise that scholars want to have a mediator selecting and authenticating web-based resources. In practice, as other aspects of this investigation have suggested, this premise is only partially true.

A10.2 Intute Features and Services.

Intute-Arts and Humanties also offers several value-added services.

Feature	Brief description	Note
AHRC Projects	Collection of AHRC funded projects that have an associated website.	New feature for Intute Arts.
Artists Index	An index of many different types of artist (which is used in its very broadest sense) for resources dedicated to 'artists' in many fields.	New feature for Intute Humanities. Work is ongoing to expand the artist roles.
Blog	Subject news from the arts and humanities fields.	New feature for Intute Arts and Humanities. Replaces previous e-newsletters.
eJournals	A collection of freely available, peer-reviewed electronic journals for a wide range of arts and humanities subjects, brought together in one place.	New feature for Intute Arts.
Harvester	A software tool which deep-searches the webpages within the websites catalogued in the Intute Arts and Humanities database, yielding many more hits. An alternative to the normal search when it provides very few results.	New feature for Intute Arts and Humanities.
Limelight	A monthly feature showcasing individual artists, topical subjects, new and noteworthy websites, or forthcoming events, exhibitions or festivals. Each feature gives information, links to related sites in the Intute Arts and Humanities database and suggestions for possible searches.	New feature for Intute Humanities.
MyIntute	Personalisation tool which includes: email alerts from Intute to stay abreast of new resources added in your field; the facility to create custom collections of resources described by Intute; the facility to export of your customised collections to use in other webpages or email selected records	New feature for Intute Arts. Builds upon the former MyHumbul service and is utilising Web 2.0 functionality.

	from Intute; the facility to create reading lists of online resources.	
News	News items and press releases from Intute, together with the Newsround service, a search interface to thousands of topical news items, sourced from hundreds of global news services; search across both news and jobs items.	Enhanced feature for Intute Arts and Humanities. Replaces former e-newsletters and external RSS news feeds.
North West Film Archive	The facility to search the North West Film Archive collection of moving images.	Migrated with the Artifact data and continues to honour an agreement with the NWFA.
On This Date	A list of Arts and Humanities-related facts and events along with suggested searches of Intute Arts and Humanities.	New feature for Intute Humanities. New content being added daily.
RSS News channels	Intute RSS news and Intute Arts and Humanities subject feeds (of the latest records added to the catalogue).	Builds upon the previous news feeds.
Seminars and events	Lists of forthcoming events and seminars in the Arts and Humanities.	New feature for Intute Arts and Humanities.
Timelines	A selection of notable events from prehistoric times to the present, broken down into separate subject areas and themes. Each timeline contains dozens of key events that have shaped the world as we know it, together with suggested searches for further information.	New feature for Intute Arts and enhanced feature for Intute Humanities.
Z39.50 target	The Intute Z39.50 target facilitates cross-searching of Intute from remote catalogues such as library systems.	

(Figure 1 Supplied by Intute-Arts and Humanities)

Comparison Between Managed Research Environment Features and Intute-Arts and Humanities

Key	
Workflow Management	
Collaboration & Communication	
Resource Discovery	

Demonstrator Features			Intute-Arts and Humanities Features	
Web- Page	**No.**	**Screenshot Featured Tools**	**Similar Features in Intute- Arts and Humanities**	**Additional Intute Arts and Humanities Features**
Home-page	1.	Desktop indexing & searching		AHRC Projects
	2.	Frequently Used Resources	MyIntute	Artists Index
Search Results	3.	Resource Usage History	MyIntute	Blog
	4.	Ranking		eJournals
	5.	Filtering		Harvester
	6.	Quality control		Limelight
Web Page Annotation	7.	Social Bookmarking		News
	8.	Copyright details/ borrowing permissions		Seminars and events
	9.	Annotations & Quote sampling	MyIntute	North West Film Archive
	10.	Referencing system		On This Date
	11.	Context sensitive searching for similar pages		Timelines
Resource Usage Information	12.	Personal usage information		Z39.50 target
	13.	Institutional usage information		
	14.	Citation history		
CV Management	15.	Linked to an institutional publication repository		
	16.	Access filtering		
	17.	Version editing		
Project Information	18.	Live chat		
	19.	Shared documents		
	20.	Archiving	MyIntute	
	21.	RSS Funding alerts	RSS News channels	
Research Partners	22.	Communication		
	23.	CVs		

(Figure 2)

A10.3 Search for Intute-AHDS Cross-listings

A search was made on the 6th of September 2006, using the search fields within Intute. The items searched were taken from the AHDS Annual Report 2004-05. In nearly all instances the exact item listed was cut and pasted into the search field. Further changes to accommodate subject ambiguity were also taken in to account, for instance only single subject terms were used. No presence of those items highlighted in red were found in the Intute search.

KEY	Found on Intute search but not linked to AHDS site
	Not found on Intute search
	Found on Intute search Linked to AHDS site

Appendix A: New Acquisitions

AHDS Archaeology
- South Yorkshire / North Derbyshire Medieval Ceramics Reference Collection
- Channel Tunnel Rail Link Archive - Phase One
- Sussex Archaeological Collections Volume 138
- HMJ Underhill Archive: The Stone Circles of Britain; The Roman Cities of Britain; Windmills
- Beads from Anglo-Saxon Graves
- Pig Measurements From Durrington Walls
- Extensive Urban Survey - Hampshire
- Facing the Palace - Excavations in front of the Roman Palace at Fishbourne
- Excavations at the Viking Barrow Cemetery at Heath Wood, Ingleby, Derbyshire
- Over 400,000 index records added to ArchSearch from the English Heritage National Archaeological Inventory
- Nailsea Glassworks study
- Predicting the Location of Hominin Sites in Africa and Asia
- Montgomeryshire Placename Database
- Trent Valley 2002: Trent Valley GeoArchaeology Bibliographic Database

AHDS History
- Credit, Class and Community: Working Class Belfast
- Deaddocs: A bibliographical index of obituaries and posthumous accounts in British Medical Journals and related sources, 1750-1850
- National Archives Class c.131: Extents on Debt, 1284-1530
- Irish Poor Law Union and Baron Boundaries, 1841-1871
- Cornish Parish Records: West Penwith and Kerrier, 1580-2002
- Scottish-registered Companies Investing Abroad, 1862-1914
- International Banking database, 1912 and 1938
- Spending and Taxation Patterns for Municipal Corporations, Local Boards and the Urban Sanitary Authorities in England and Wales selected years 1868-1888
- The Number of Voters on the Burgess Roll in Municipal Boroughs in England and Wales, 1852, 1865, 1871 and 1884
- Charity in the Southern United States, 1800-1860
- Foxe's Book of Martyrs Variorum Edition
- Digital Library of Historical Directories,

AHDS Literature, Languages and Linguistics
- Early Stuart Libels: English Language and Literature

- French Learner Language Oral Corpora (FLLOC): Linguistics
- Morphosyntactic Typology of Romani database: Linguistics
- Newcastle Electronic Corpus of Tyneside English: Linguistics
- Anglo-Saxon Charter Boundaries material: Medieval and Modern History
- Biblia Sacra Juxta Vulgatam Clementinam: Religious Studies and Theology
- Bibliography of Hausa Popular Literature 1987-2002: Modern Languages
- Sheffield Corpus of Chinese, Xiaoling Hu: Linguistics
- The Auchinleck Manuscript: English Language and Literature
- The Chambers-Rostand Corpus of Journalistic French: Linguistics
- Agamemnon: Classics

AHDS Performing Arts

- Dance Data On-Line
- Embodying Ambiguties: intertextual plays within and between space-time philosophies and the performing body.

AHDS Visual Arts

- Courtauld Institute of Art, Corpus of Romanesque Sculpture in Britain and Ireland
- Crafts Study Centre L & T module, Calligraphy: an education in letter form
- Crafts Study Centre L & T module, Hand-blockprinted Textiles: Phyllis Barron and Dorothy Larcher
- Crafts Study Centre L & T module, Pioneers and their practice: a reference guide
- Crafts Study Centre L & T module, The Life and Work of Bernard Leach
- Crafts Study Centre L & T module, Thirteen Weavers
- De Montfort University, Photographic Exhibitions in Britain 1839 -1865
- Design Council Archive L & T module, Art for social spaces: Public sculpture and urban regeneration in post-war Britain
- Goldsmiths College, University of London, Constance Howard Resource and Research Centre in Textiles: Material Collection - deposit 1
- Imperial War Museum, Imperial War Museum: Posters of Conflict - deposit 1
- London College of Fashion, London College of Fashion: The Woolmark Company - deposit 2
- Manchester Metropolitan University (MMU) , Design Council Slide Collection - RSLP deposit
- South Asian Diaspora Literature and Arts Archive (SALIDAA)
-      SALIDAA, Akademi Collection
- SALIDAA, Amal Ghosh Collection
- SALIDAA, Tamasha Theatre Company
- SALIDAA, Tara Arts Collection

Arts and Humanities Data Service. (2005) "AHDS Annual Report 2004-2005."
http://ahds.ac.uk/about/reports-and-policies/ahds-annual-report-2004-05.doc

(Figure 3)

Appendix A11 RePAH Project Activities and Outputs

Greengrass: presentation to Arts and Humanities Research Council ICT Strategy Review Projects Meeting	December 2005
Bryson: RePAH Poster Presentation to Director of the AHRC, Professor Philip Esler	May 2006
Bryson(2006) "Managing Web-based Information in an Arts and Humanities Research Environment," in *Portals: People, Processes and Technology*, ed. Andrew Cox. Oxford: Facet	June 2006
Brown, S, Bryson, J, Greengrass, M, and Ross, R (2006) 'AHDS Review and User Survey' A report to the Joint Information Systems Committee (JISC) and the Arts and Humanities Research Council (AHRC)	July 2006
Brown, Greengrass, Ross, Gerrard: Digital Resources in the Humanities and Arts Conference, Devon, UK	September 2006