Research portals in the arts and humanities

Stephen Brown
Knowledge Media Design, De Montfort University, The Gateway, Leicester LE1 9BH, UK

Mark Greengrass
Humanities Research Institute, Sheffield University, 34 Gell Street, Sheffield S3 7QY, UK

Abstract

There has been dramatic growth in information communication technologies (ICT) infrastructure for the arts and humanities research community in recent years in the UK and elsewhere. No domain-wide survey of how researchers are using ICT and what they perceive their future needs to be has been undertaken previously and consequently what is needed in terms of a generic ICT infrastructure to support arts and humanities research is not well understood. The RePAH (Research Portals for the Arts and Humanities) Project is an AHRC funded study into the user needs for information portals to support research in the Arts and Humanities. It uses a combination of questionnaires, focus groups and Delphi opinion gathering, combined with server log-analysis data, to identify users' information discovery strategies, Internet usage patterns, awareness and attitudes towards current services and technologies and responses to what future portal developments can deliver. Responsibility for funding this kind of infrastructure is split between a number of different agencies. This makes sector-wide information gathering for strategic planning and development difficult. The results of this study may help providers to understand where the priorities lie for the arts and humanities research community and help users to appreciate some of the possibilities within their grasp.

Correspondence:
Stephen Brown,
Knowledge Media Design,
De Montfort University,
The Gateway, Leicester LE1
9BH, UK.
E-mail: sbrown@dmu.ac.uk

1. Introduction

The infrastructure of academic scholarship in the arts and humanities comprises the institutional fabric of libraries, archives, museums, research centres, etc. plus the tools of scholarship—bibliographies, searching aids, concordances and editions, journals and academic presses—that make information accessible. The Internet makes possible the development of an equivalent digital, online, infrastructure for academic scholarship and the UK has been at the forefront over the past decade in developing an e-infrastructure of digital repositories and subject gateways to support Arts and Humanities research.

In 1995 the Joint Information Systems Committee (JISC) began operating a number of subject gateways including HISTORY (a gateway for History) and ADAM (Art, Design, Architecture and Media information gateway). By 1999 the JISC

Literary and Linguistic Computing, Vol. 25, No. 1, 2010.
doi:10.1093/llc/fqp032 Advance Access published on 5 October 2009

subject 'gateways' had evolved into broader subject web-based Resource Discovery Network (RDN) 'hubs' including Humbul, that reviewed and catalogued websites of interest to academics and students in the humanities (Pinfield and Dempsey, 2001; Hiom, 2006) and later, Artifact, a hub to support the visual and performing arts. In 2006 these separate RDN hubs evolved further into an integrated network called Intute[1] that 'allows access to both subject-specific and cross-subject resources, all of which have been evaluated for their quality and relevance'.[2] Intute: Arts and Humanities is part funded by the AHRC and the total cost of Humbul and its equivalents between 1995 and 2009 is estimated at £1.76 million.

In addition to developing the RDN, the JISC joined forces with Kings College London in 1995 to establish the Arts and Humanities Data Service (AHDS). AHDS, launched in 1996, was established to collect and preserve as well as promote electronic materials resulting from research and teaching in the arts and humanities. In 1998 the then Arts and Humanities Research Board (AHRB) took over funding AHDS from Kings College London and its role as a curator of electronically created materials was substantially enhanced by the decision of the AHRB in 1999 to require funded projects which produced electronic content to deposit it with the relevant AHDS service. This resulted in a sharp rise in acquisitions in 2004–05,[3] amounting to 1,225 collections by 2006. Total funding for the AHDS between 1995 and 2006 was just under £8.7 million (Greengrass, 2006). In May 2007 the AHRC announced a decision to withdraw funding for AHDS as of March 2008.

Other significant developments have been the development of the JISC Collections,[4] a procurement scheme to purchase access to valuable data sets on behalf of the wider research community; the Arts and Humanities Research Council's Resource Enhancement Scheme (discontinued in 2006) which funded digitization of primary research resources to the tune of amost £40 million, the AHRC funded information communication technologies (ICT) Methods Network[5] which ran from 1 April 2005 to 31 March 2008 to promote the use and exchange of advanced ICT methods in Arts and Humanities research; and the AHRC/EPSRC/JISC funded Arts and Humanities e-Science Support Centre (AHeSSC),[6] set up in association with AHDS as part of the AHRC-JISC e-Science Initiative, a £1.8 million national programme to promote and develop e-Science in the Arts and Humanities; and the AHRC ICT in Arts and Humanities Research programme funded at £3.8 million from 2003–08 to build national capacity in the use of ICT for arts and humanities research and to advise the AHRC on matters of ICT strategy. Most recently we have seen the establishment of arts-humanities.net[7] hosted by the Kings College London Centre for e-Research and the Network of Expert Centres: 'a collaboration of centres with expertise in digital arts and humanities, in the sense of data creation, curation, preservation, management (including rights and legal issues), access and dissemination, and methodologies of data use and re-use.'[8]

Exponential growth in online resources created by a wide variety of different content-creators and contractors is fostering a growing level of expectation as to the availability of research materials in digital form across all the Arts and Humanities and affecting the way research in these domains is conducted (Unsworth, 2003). Over time we have seen a shift in emphasis from investment in access to resources to concerns about use of these resources, skill levels and attitudes towards use of ICT in arts and humanities research and a consequent closure of some services and the emergence/resurgence of others. However, the arts and humanities constitute a wide spread of disciplines with an equally broad spread of research traditions and it can not be assumed that innovations in one discipline necessarily meet the requirements of others.

It was against this background that around 2004–05 the JISC and AHRC commissioned a number of studies looking at current ICT skill levels among Arts and Humanities researchers, use of available online research data sources, attitudes towards ICT in research and reactions to possible future scenarios of ICT-based research environments. This article reports on the findings from one of these, namely the Research Portals in the Arts and Humanities project (RePAH), carried out by the Humanities

Research Institute, Sheffield University and Knowledge Media Design at De Montfort University and funded by the AHRC ICT in Arts and Humanities Programme. This was an information-gathering project aimed at discovering user-behaviour and Arts and Humanities researchers' needs and expectations with respect to online research portals. For full details of the project and the final report, see http://repah.dmu.ac.uk.

While an information gateway places the emphasis on providing links to distributed sites of information, a portal places the emphasis on federating distributed sites of information. According to the JISC[9]:

> 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 collates 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.

Hitherto there has been no sector-wide comparative study to ascertain how researchers are using ICT and what they perceive their future needs to be. Consequently what is needed in terms of an ICT infrastructure to support Arts and Humanities research is not well understood. Are there, for example, significant differences in the ways in which researchers from different disciplines use ICT in their research? Are some domains more technically advanced than others? How widespread is ICT-based research across the sector? Can a single portal concept meet the needs of the whole community?

Responsibility for funding this kind of infrastructure is split between a number of different agencies. This makes sector-wide information gathering for strategic planning and development difficult. The results of this study may help providers to understand where the priorities lie for the Arts and Humanities research community and help users to appreciate some of the possibilities within their grasp.

2. Previous work in the area

The discussion of user requirements for ICT in arts and humanities research purposes substantially predates the emergence of the portal concept in the mid-1990s (Stone, 1982). However, notwithstanding this we have located only a small number of previous studies focused on either specific subsets our our broad target group, on particular technologies or features.

> A 1996 preliminary survey of user information needs and search needs was undertaken by Alison Ferry to inform the design of the ADAM gateway in art, design, architecture and media studies (Ferry, 1997). It was based on 723 completed responses to a distributed questionnaire.
>
> A user-needs survey was conducted by the Visual Arts Data Service in December 1997-February 1998, based on a paper and online questionnaire, to which there were 107 responses (Grout and Rymer, 1998)
>
> 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 (Greenstein, 1998).
>
> A user-needs analysis for electronic information gateway provision in archaeology, was undertaken in the spring and summer of 1998 for HEIRNET. It was based on 3,000 questionnaires, mailed to archaeologists and followed up by a smaller number of structured interviews conducted in July 1998. (Condron, Richards, Robinson and Wise, 1999).
>
> A preliminary user-survey (Guy, 2003) in which user scenarios were developed 'in-house' by some RDN hubs to help to scope

the requirements for their planned alerting and user-profiling services.

A more general survey of portal functionality was undertaken by ALTIS, information scientists and specialists at the University of Birmingham and a part of the RDN (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. 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.

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)

The Nielson-Norman Group Report of 2005 (Goodwin, Schwartz and Nielson, 2005). This latter 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].

How, if at all, most of these evaluations fed through to modifications in the design, presentation and functionality of the services in question is unclear. A notable exception is the HEIRNET study which was broad-ranging, strategic and undoubtedly had a significant impact in developing service provision in that specific area. HEIRNET subsequently undertook a further user-evaluation survey in 2002, commissioned from the Cultural Heritage Consortium (2002). This was reinforced by a further 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 Council for British 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.

3. Methods

The notion of 'users' is clearly not being applied consistently across all these previous studies. In some cases it refers to existing online service users, in others attempts have been made to include broader sections of the community. For the purposes of this study users have been defined very broadly as the 'arts and humanities research community'as encompassed by the (then) eight panel profile of the AHRC mapped to the RAE subject panels. We estimate it as around 50,000–60,000 active practitioners, comprising Faculty and Independent Researchers, Postgraduate, Postdoctoral and Research Assistants (Brown *et al.*, 2006, p. 7). Figure 1 shows the range of players and activities covered.

RePAH had a requirement to discover:

- Information about users' information discovery strategies and internet usage, in particular patterns of recent user-activity in relation to the JISC and AHRC funded services.
- Information about users' awareness and attitudes with respect to available online services and tools, including such repositories, gateways and portals as exist and their needs and expectations with respect to possible future developments.

This combination of actual behaviour and less tangible attitudes and awareness required a corresponding combination of data collection methods. We chose to use five main methods:

(1) Questionnaire surveys
(2) Focus groups and interviews
(3) Server log analysis
(4) Delphi forecasting
(5) User trials

3.1. Questionnaire surveys

An online questionnaire on the project website was used to gather information from users about their

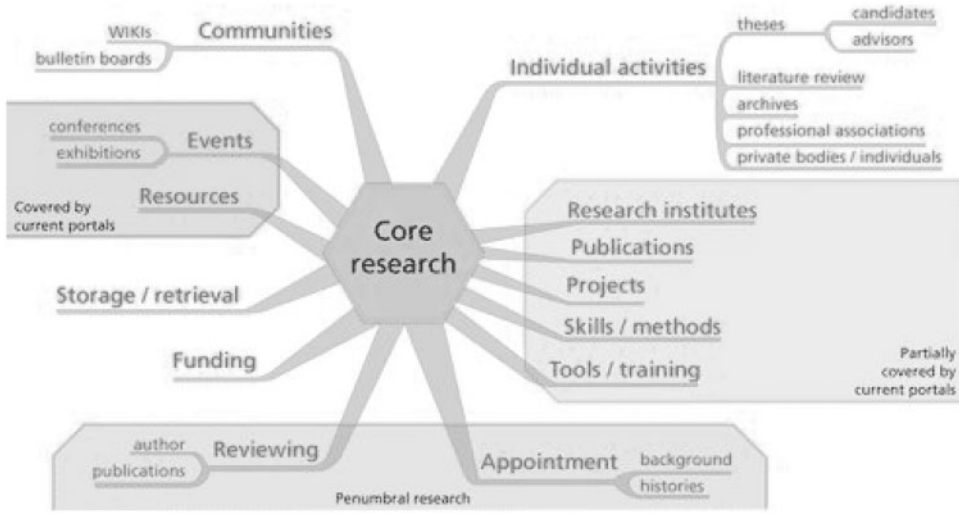

Fig. 1 'Road Map' of Arts and Humanities Research

research backgrounds, Internet use habits and attitudes towards online resources. The questionnaire was linked to from a number of other sites, in particular AHDS and Humbul and open from 1 December 2005 to 30 April 06. 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 across the arts and humanities spectrum. One hundred and forty-nine people returned completed questionnaires, representing a good spread of researchers from different domains. Respondents were frequent Web users. Eighty-nine percent used the web on a daily basis and 77% had been using the Web for 5 years or more. Most respondents regarded themselves as multidisciplinary. Only 31% said they worked within a single domain and one individual claimed to work across eight! But overall the respondents were distributed fairly evenly across the research domains covered by the AHRC.

3.2. Focus groups

Focus groups combine elements of interviewing and participant observation. The group is presented with a series of questions by a facilitator/observer. The advantage of focus groups over interviews is the interaction between participants which can generate data and insights that would be unlikely to emerge from one-to-one interviews. We used focus groups to capture qualitative data on topics indicated by the online questionnaire results and the server log analysis data. Five focus groups were drawn from the University of Sheffield departments of archaeology, history, biblical studies, music, and information studies. These were supplemented by interviews with individual scholars in the arts from De Montfort University, including lecturers in creative technology, music studies, digital imaging, visual arts and holography. Each of the focus groups consisted of three to seven participants from among full-time contract researchers and lecturing staff. One of the focus groups did invite postgraduates, who were able to contribute their experiences of Web-based work required as part of their research curriculum.

3.3. Server log analysis

Web-server logs record simple traffic statistics and data such as number of page requests per month and originating addresses of page requests and can be analysed to reveal numbers of registered users, numbers of visitors, patterns of behavior, trajectories through Web sites and popularity of

individual pages or sections. In principle server logs provide objective data about actual behaviour which can be used to triangulate information provided through the questionnaires and focus groups. In practice however, there are often problems with the interpretation of such data (Warwick *et al.*, 2008). Server logs from AHDS servers, Humbul and Artifact were analysed by another AHRC project, LAIRAH, who kindly shared their results. The LAIRAH study has been reported elsewhere (Warwick *et al.*, 2008), so this aspect of the project is not reported in detail here.

3.4. Delphi forecasting

Delphi is 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 (Weaver, 1971). It measures the degree of consensus among the panel regarding future possibilities where the decisive factors are subjective, and not knowledge-based. It possesses the benefits of group decision making while insulating the process from the limitations of group or peer pressure and overly dominant individuals. We used Delphi to filter the ideas that came out of the focus groups by asking participants to rank them in terms of which would be the most important for their own future research. The exercise was conducted in three rounds between February 2006 and March 2006 with a panel of 109 researchers, of whom 106 participated. The panel comprised all members of the focus groups, other experts known to the researchers, plus respondents to the online survey questionnaire that had agreed to being contacted for further information.

3.5. User trials

User trials are a technique for gaining user responses to design ideas, working from mock-ups or simulations (Hix and Hartson, 1993). What users think of an abstract idea can be different from how they react to a physical manifestation of that same idea (Rettig, 1993). So we used user trials to cross-check the results of the focus groups, questionnaire and Delphi exercise by presenting users with a range of possible future portal features and tools to determine which would be most valuable in a virtual arts and humanities research environment. Presentations were made with Web-style screen mock-ups supplemented with scenarios of the uses to which the various tools presented could be put. The proposed features were:

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

The demonstrators were designed to be essentially modular in nature to allow extension and personalisation. As a result, thy do not cover all the potential functionality of a system of this sort. Instead, the following were highlighted:

- 'The system homepage': what the researcher would see when they logged on using their Shibboleth or other user authenticated account.
- 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.
- 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.
- 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 researcher's homepage (Figure 2) provides a general overview of the functionality of the entire system and a way into all the different areas. 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 drop-down that allows the user to choose which of a list of search engines the term should be searched for within. The local university's library OPAC might be one of the suggested search locations, but the institution might make new options available, and the researcher might also be able to customise this list with their favourite search engines.

It is worth noting the blank space below the tool bar, which is a kind of 'activity bar' which will change to include functionality relevant to the different parts of the system when the researcher is within them.

'Your documents' is inspired heavily by the Google desktop application, which indexes and searches documents on your local hard drive. '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. If the local research portal server were in communication with a central HE 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.

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.

Figure 3 shows the first few results of a search using Google Scholar, undertaken with the 'search bar' on the homepage. 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. 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.

Figure 4 shows a web page visited by a researcher who is logged into the system. This functionality

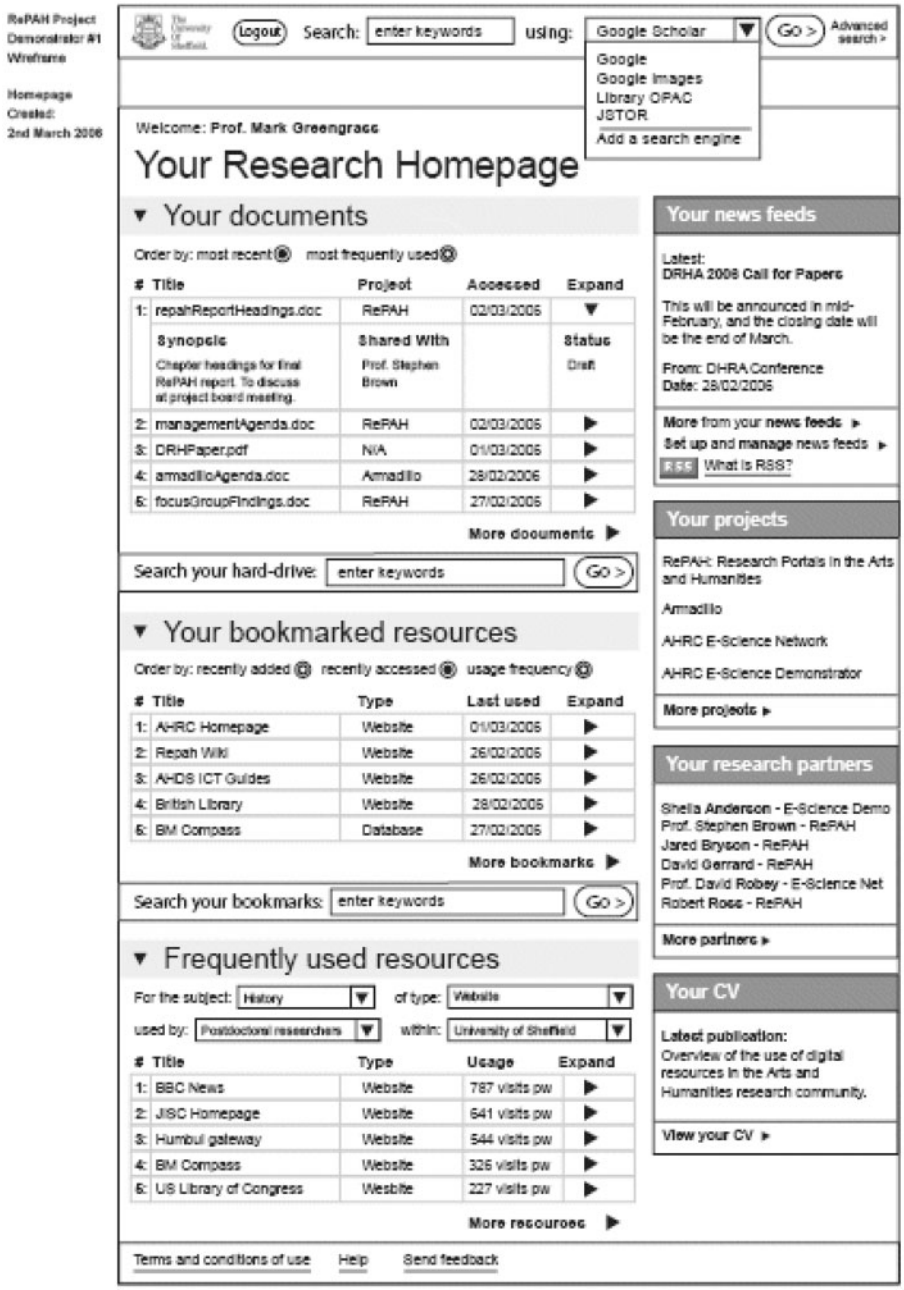

Fig. 2 The researcher's homepage

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.

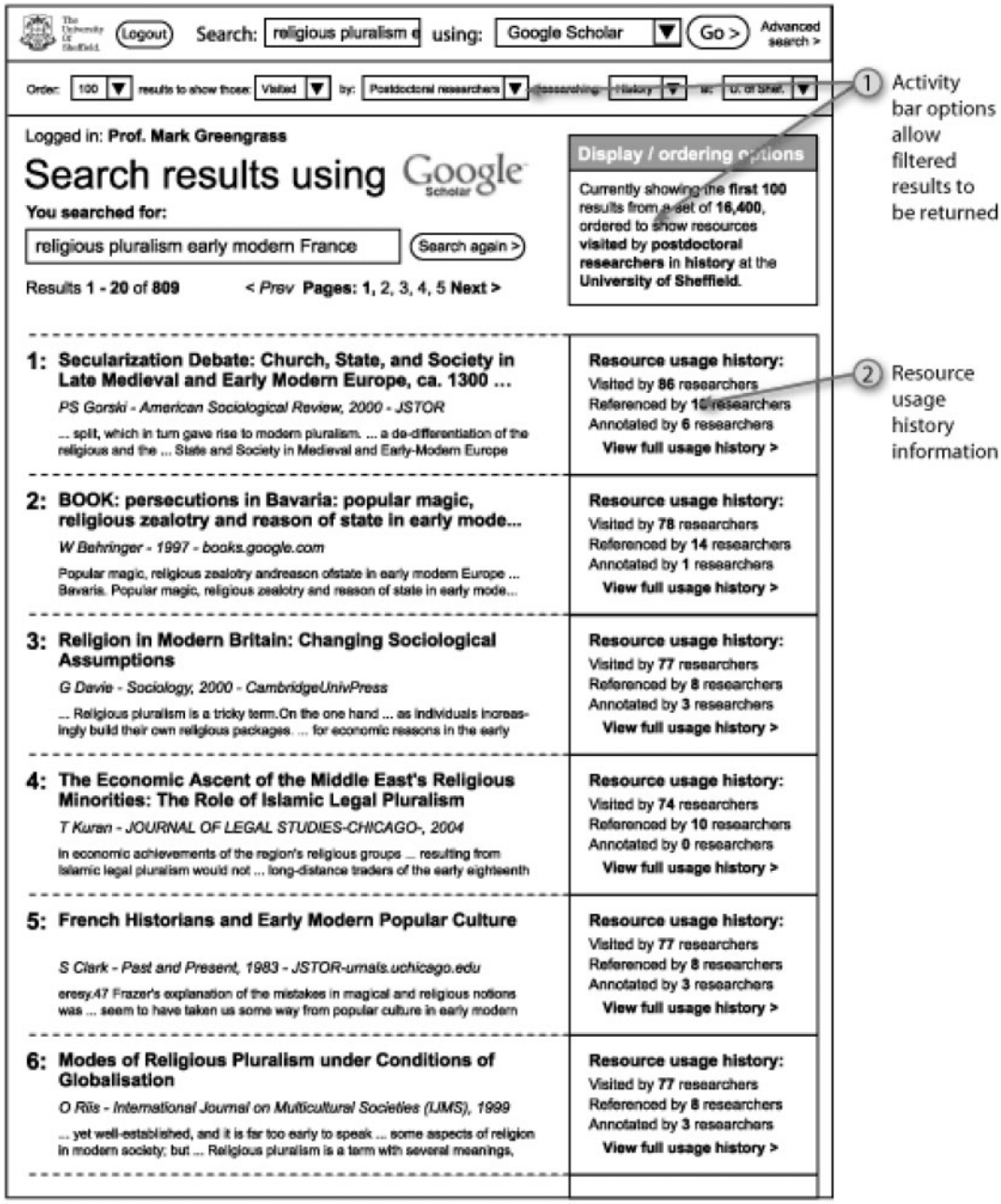

Fig. 3 Search results page

(1) A link to the full 'usage history' for this web page.
(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

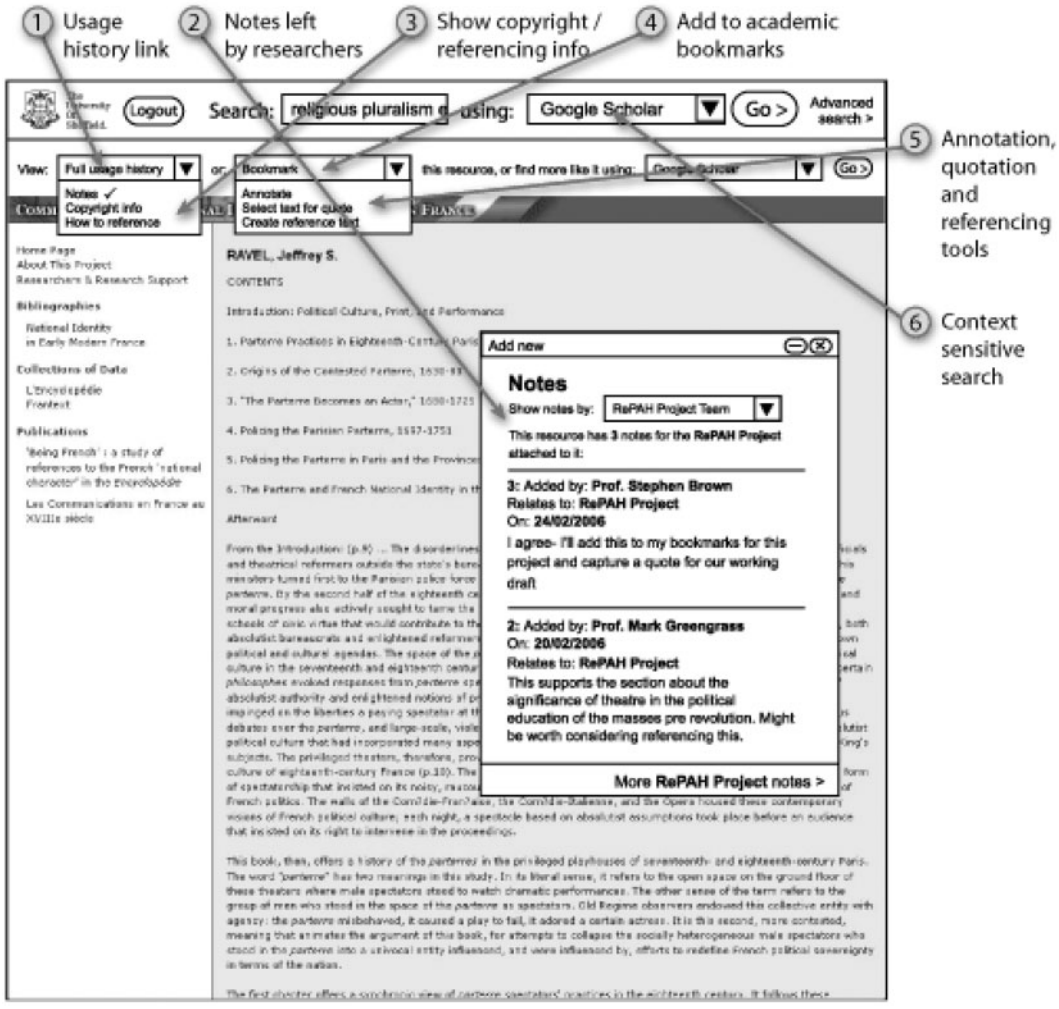

Fig. 4 Web page annotation

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, 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.

Figure 5 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, or by

looking at a bookmark, or by browsing through the 'frequently used resources' selected by other researchers from the local institution or UK HE as a whole. It would also be possible for the researcher to view the usage histories of their own publications via their CV page. 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.

Figure 6 shows 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, 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. Expanded bookmarks also show the latest note added about a resource, and link to documents and publications in which the researcher had referenced the resource. 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).

Figure 7 shows the state of the screen as the researcher administers their own CV, which means that 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). The activity bar allows the researcher to view and roll back to previously saved versions of their CV.

Figure 8 shows all the information about a particular project (in this case the RePAH project itself).

(1) The members of the 'project team': clicking their names would display their CV.
(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.
(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.
(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.
(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 final layout (Figure 9) shows the researcher's project partners, which they would use in a similar fashion to their bookmarks. From this page they could view all the documents shared with their partners; link to the projects they were working with their partners upon; read email sent to and received from partners; contact and view the CVs of all the partners they were currently working with; view all bookmarked resources recommended by partners.

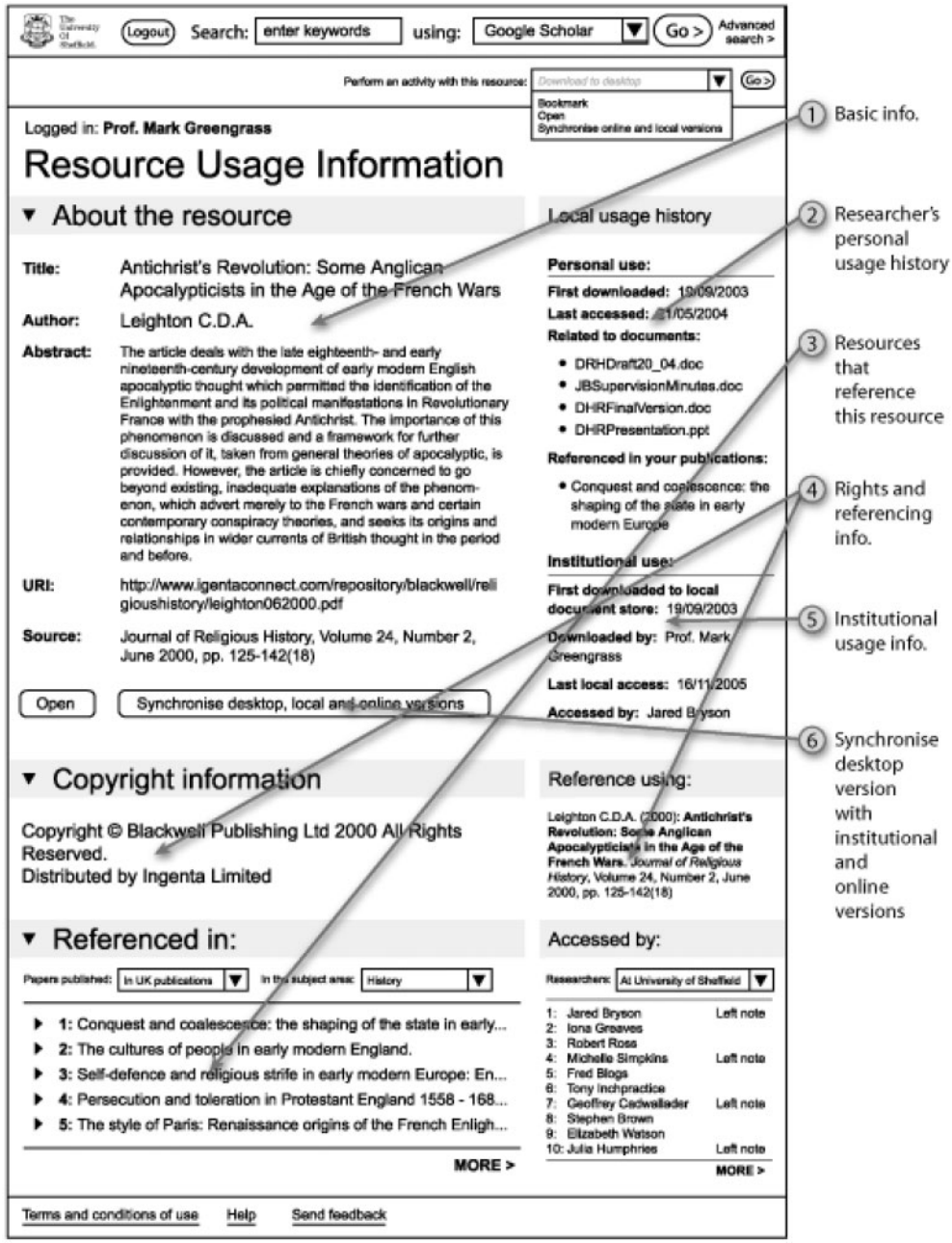

Fig. 5 Resource usage information

4. Results

4.1. Pull versus push

Over 60% of respondents to our online questionnaire regarded digital resources as 'essential' to their research. These resources were used 'extensively' by a majority of our respondents and 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. But respondents saw the Web primarily as a source of data that they could pull down when needed rather than as a repository into which they could push their own data. Collection and analysis of information was seen as central to the work of over 50% of the respondents, whilst data storage and archiving was not given such a high level of significance.

4.2. Advantages

Focus group participants identified three ways in which Web Information resources had affected their work:

(1) Speed and efficiency: Accessing texts, images and artifacts within their particular subject speciality is enormously more convenient than visiting physical library collections—even if they are based at the researcher's own institution. The widespread availability of Web access enables texts to be reviewed from a laptop computer nearly anywhere and the ability to search those texts by key words has increased the efficiency of information discovery and knowledge creation.
(2) Timeliness: The printed works found in institutional libraries are often burdened with a time-lag not found to the same extent among the digital resources.
(3) New ways of working: Several times during the sessions researchers referred to using the Web as a tool to help them think—clarifying ideas and discovering new ways of approaching various research problems.

4.3. Sources

The questionnaire responses revealed that across the full spectrum of arts and humanities research a very broad range of resources are accessed, with little commonality between different disciplines. Not surprisingly the most commonly cited resources were generic such as University library services, for example COPAC or associated access permissions to resources (such as journals) gained from it, as shown in Fig. 10. The next most frequently quoted resource was Google and its attendant functions such as Google scholar or Google Images with JSTOR and AHDS services the next most quoted. 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. But by far the largest category of resource cited was 'other' at 63% (Fig. 10). Other resources were those mentioned by just one or two researchers. This clearly has implications regarding the possibility of one solution meeting the needs of all researchers.

The popularity of search engines was confirmed by the server log data which showed that around 60% of the traffic coming into the Humbul RDN hub (now Intute Arts and Humanities) and up to 40% of AHDS traffic entered via a search engine. Focus group participants who reported that the Google search engine was their preferred application for accessing the Web explained this was primarily for its simplicity and ease of use.

4.4. Shortcomings identified

Most focus group participants reported that they were satisfied with the resources they currently used although they would prefer greater access to their subject's literature, especially journals. However, the quality and quantity of search returns were regarded as serious problems: too many hits yielding non-relevant results. Therefore, what was wanted were tools for aggregating data for searching and analysis and better quality control and ranking of results. They were suspicious of the ranking of the hits returned by search engines, but were equally overwhelmed by the information redundancy which accompanies search-engine retrieval on

internet materials. Our users wanted to have assurances of quality. This emerged in the first focus groups. 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. In addition, users were 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.

Restrictions on access and use related to copyright and intellectual property rights were also seen as a growing concern. Any technologies that might facilitate open access to these resources should be a priority. Other useful features identified were online collaboration tools, grid connections and services, personalisation and bookmarking tools, desktop video conferencing, peer reviewing facilites, and pushed alerts, for example for news on conferences, papers and funding.

4.5. Researchers' views on portal features

Respondents were positive about the potential that the proposed resource management tools offered but the overarching message that came out of the user trials was they wanted simple tools that required little or no input of time or personal engagement.

Resource discovery tools that provided greater control over web-based resources were highly valued by researchers. The ability to filter the quality of search results and to search multiple databases were at the top of all responses. Journal articles and online bibliographical resources were 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 own 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.

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 were also highly valued.

Automatic information-harvesting tools were highly valued when applied to digital content to which users wanted access. 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:

(1) automated monitoring of electronic resource usage by research practitioners (to assist in shaping user-needs for the future)
(2) 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.

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.

Advanced communication tools were not valued highly. Users were satisfied with existing communication systems, particularly email. Real-time 'chat' and desktop video-conferencing ranked consistently among the lowest of all tools proposed.

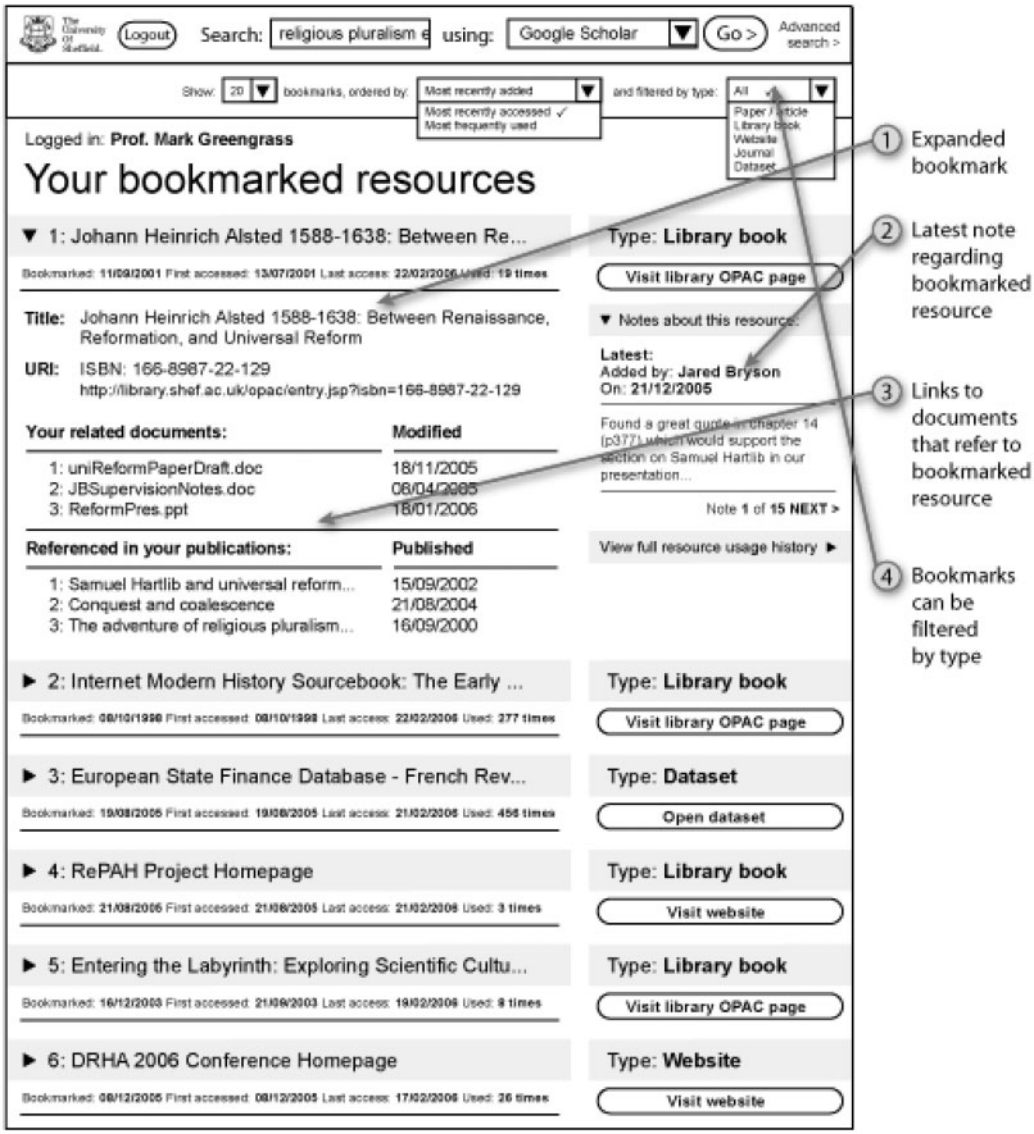

Fig. 6 Bookmark management system

5. Discussion and conclusions

Our investigation revealed the importance of digital resources for Arts and Humanities research. However, despite its impact on research, ICT has not fed through to the habits and procedures for personal digital data archiving, and has not yet had a substantial impact on the means of scholarly communication in the arts and humanities. In short, it has not yet profoundly influenced the way in which arts and humanities publication is conceived. The stereotype of the lone scholar who publishes via a printed journal article is still the dominant model across the community. So features relating to archiving and storage of data, online annotations and collaborative working are not priorities for future portal development.

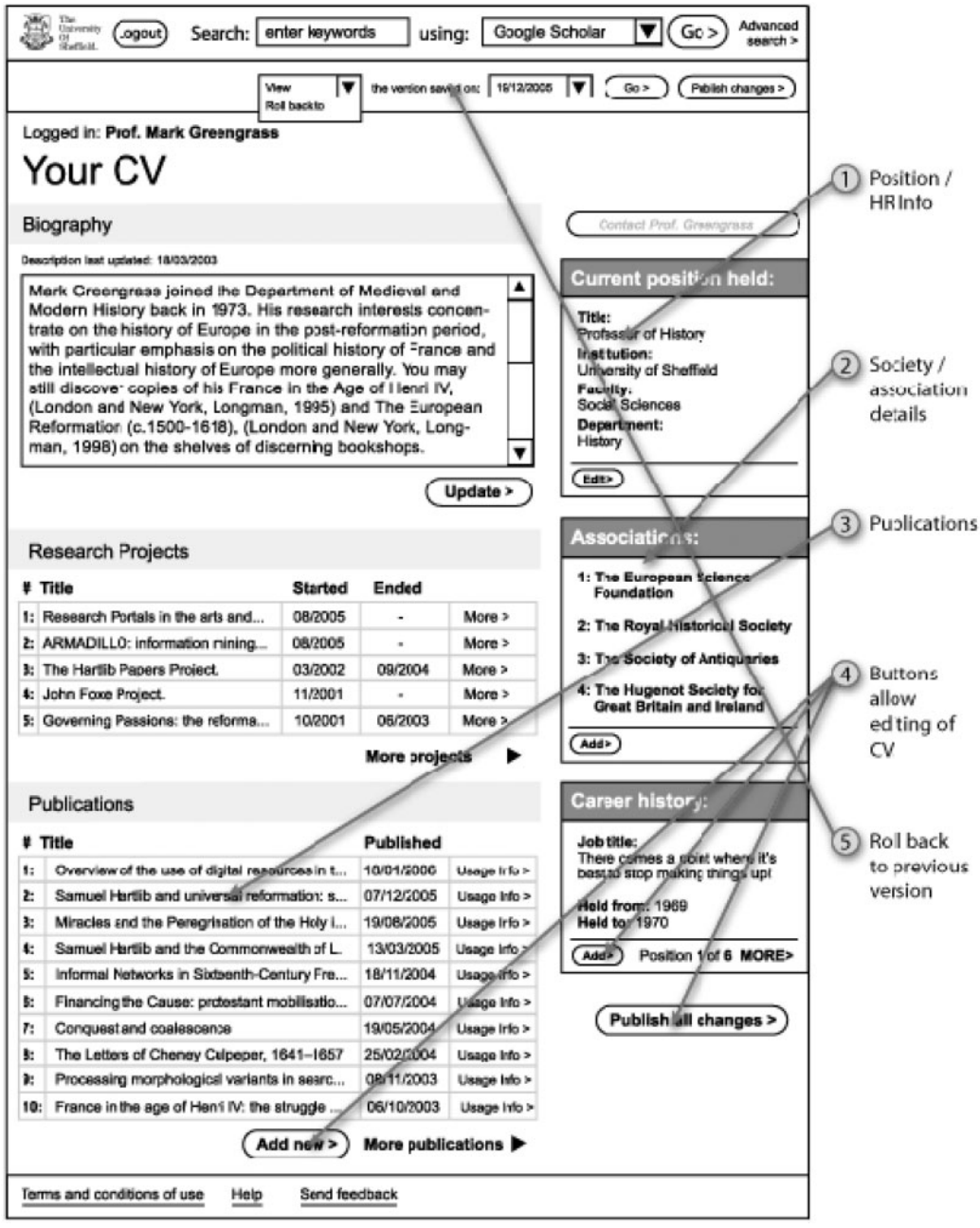

Fig. 7 Researcher's online CV

Resistance to automated monitoring of individual activity in order to enhance the performance of the system for the community as a whole is a further indication of the individualistic nature of the community. These suggestions raised issues for our users about the potential infringement of personal privacy. They challenged the predominantly individualistic scholarly culture. There was a

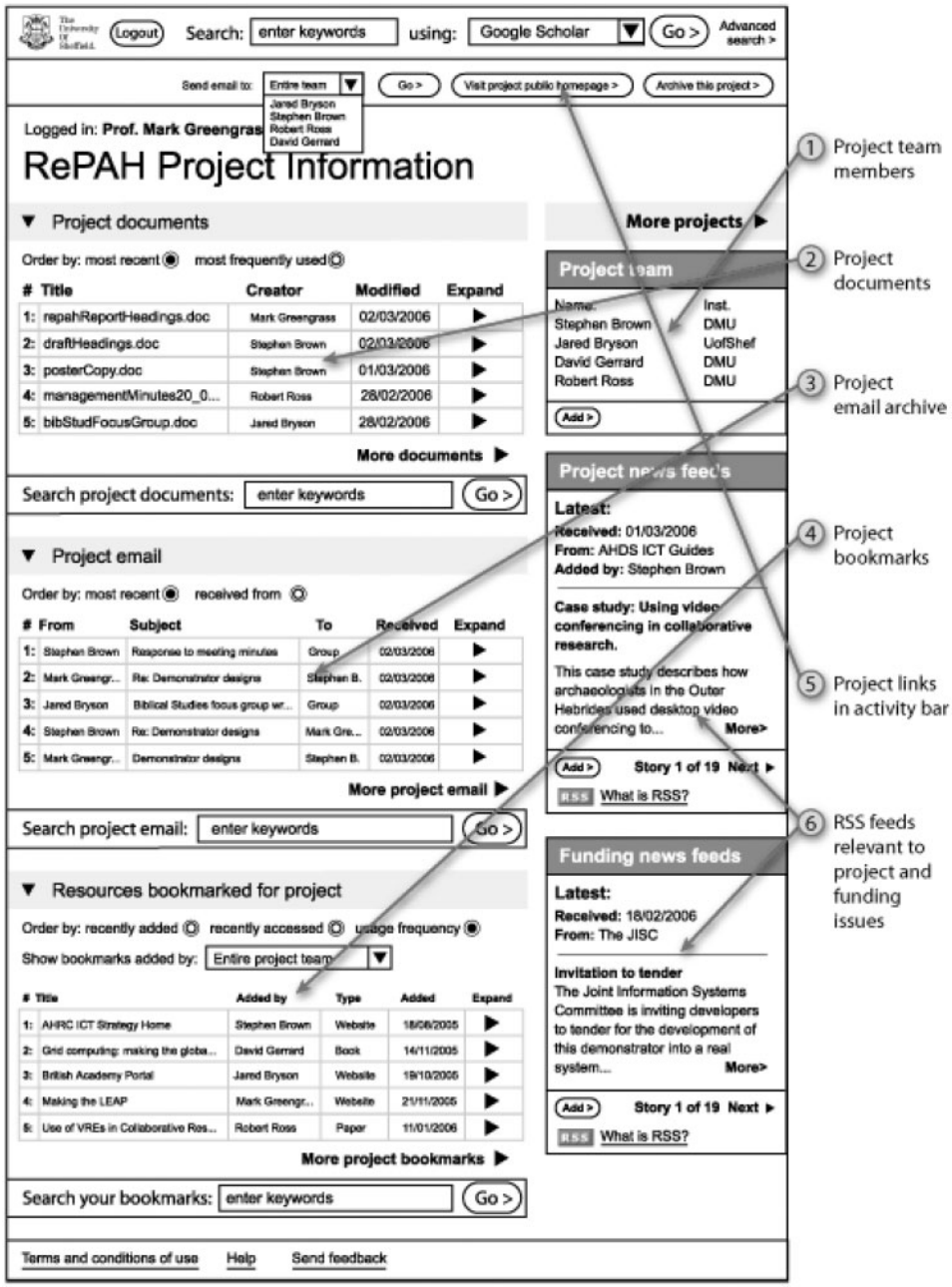

Fig. 8 Project information page

concern, particularly marked among early-career academics, about the possible abuse of such information. Since 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 judgment about their relative standing in their field, the concerns raised here suggest a lack of awareness about the extent to which actions are already monitored and recorded.

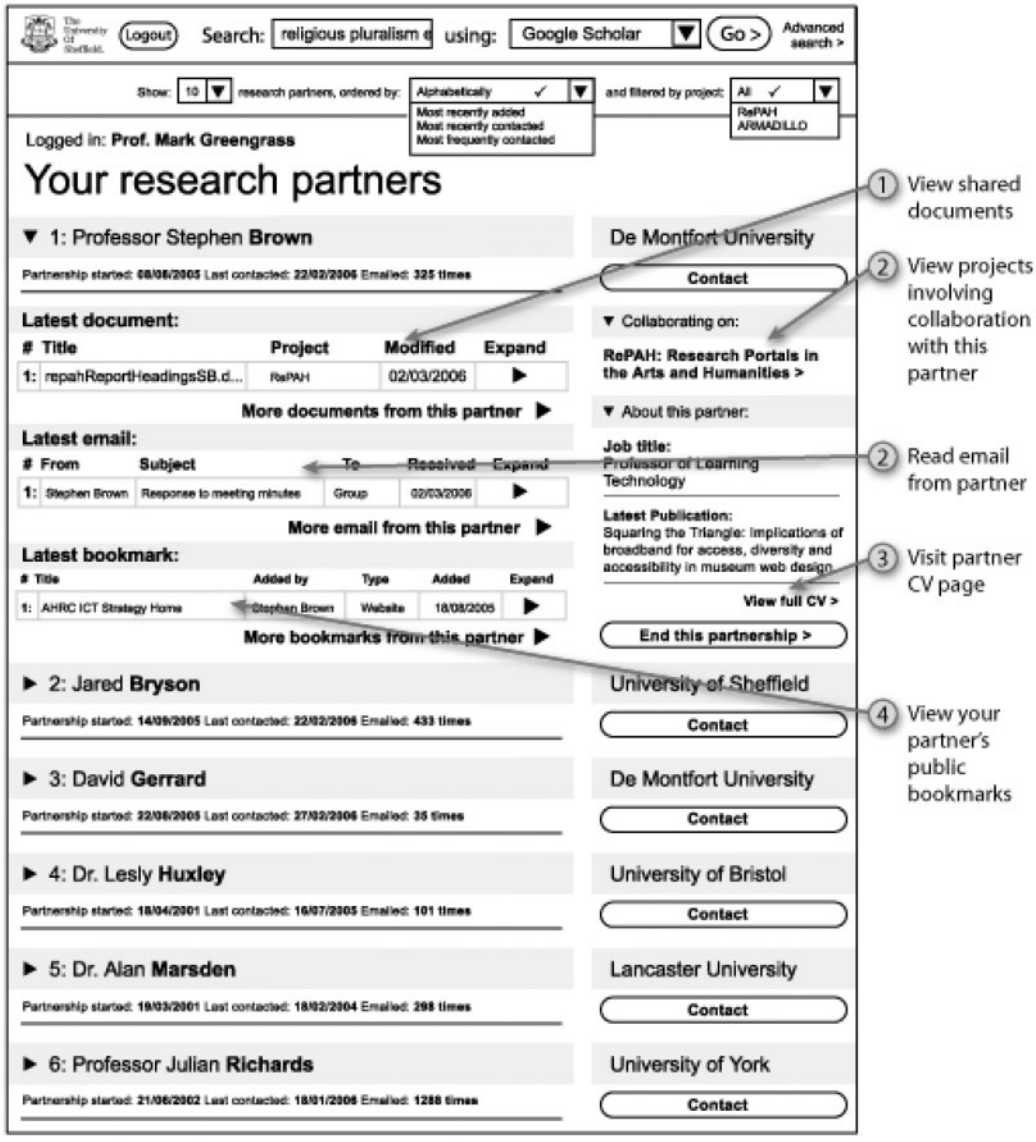

Fig. 9 Research partner page

When this is coupled with the strongly expressed preference for simple tools that require little or no learning and their expressions of frustration at the lack of sophistication of search engines (a frustration that was often a function of their lack of familiarity, or perhaps understanding, of Boolean search parameters permitted in Google's advanced search facilities), a picture emerges of researchers with relatively limited technical skills. Our focus group participants reported levels of formal initiation or training in the digital resources that they used varying from little to none. The implication here is clearly that future portal developments should assume only a very basic level of ICT competence.

Although researchers were nervous about their own information being harvested, automatic information-harvesting tools were highly valued when applied to digital content to which users

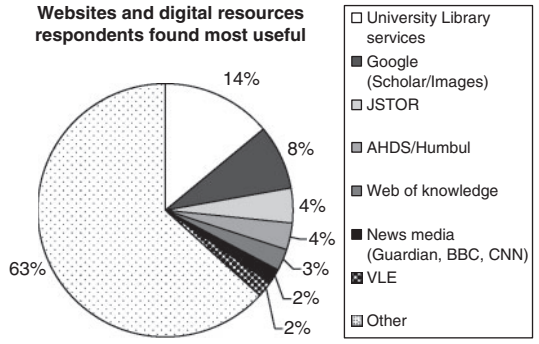

Fig. 10 Respondents most quoted digital resources

wanted access. The issue of 'access' runs through all our results. 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. Arts and Humanities researchers seem to work with a very broad range of digital libraries, bibliographical tools, encyclopaedia, dictionaries, and other online materials. Their needs are extensive, and often indeed broad-ranging and the equally extensive online bookmarked resources referred to by our respondents present both a challenge and a steer for future portal developers. The breadth of resources required to service the needs of such a heterogeneous community is unlikely to be encompassed by any single repository, or even a small cluster of major repositories. An access portal therefore needs to be 'customisable' to create links to and feeds from valued and commonly used sources.

Internet search engines emerge from this study as immensely useful digital resource discovery tools. Their simplicity and speed appealed to our users, for whom a key determinant in their cost-benefit analysis of resource discovery tools was whether they saved, rather than cost them time. That said, our users were also often aware of the limitations of their internet search engine of choice. Again this has far-reaching design implications, suggesting perhaps that search-based portals are likely to be more successful than content-based ones but the user interface needs to tread a careful path between simplicity and sophistication. Our respondents were also concerned about the quality and relevance of search results and who was undertaking the quality assurance. Future portals should therefore offer users a role in the process of quality assurance. This however will be problematic while there is continued resistance to automated monitoring of individual activity and open publication of online CVs as users need to be able to assess the standing or reliability of judgements made by their peers.

Finally, the sophisticated, lateral research networks in the Arts and Humanities seem adequately served by the current range of email, bulletin boards, and blogs. Additional communication features are not apparently a priority for portal developments.

In conclusion, there is an appetite for and a need among arts and humanities researchers for more sophisticated online research infrastructures that combine easy open access to content with simple to use search engines that nevertheless are capable of delivering relevant, quality assured results. The interface for such future portal developments should assume only a very basic level of ICT competence but should be customisable to create links to and feeds from valued and commonly used sources.

Arts and humanities researchers do not in general want tools to support collaborative working, online archiving or publishing and they are suspicious of tracking systems depite their potential to underpin powerful quality assurance systems. It is apparent that for many current researchers there is some considerable effort needed to appreciate what is already available and some of the possibilities within their grasp.

Acknowledgements

The authors would like to thank the AHRC ICT in Arts and Humanities Research programme funding this project. Thanks are due also to the LAIRAH project for sharing their Server log data and analysis with us. Finally, we would like to thank other members of the project team: Robb Ross, Dave Gerrard, Jared Bryson, members of the project advisory board, the gatekeepers of the various conduits that we used to communicate with the research

community and all the participants who took time to engage with us.

Appendix

Here are some brief definitions of some of the terms used in this article. For a comprehensive treatment of this topic see the AHDS ICT guides site http://ahds.ac.uk/ictguides/developed by the corresponding author in collaboration with the AHDS.

Advanced communication tools: go beyond basic email to include voice over IP telephony (VoIP) such as Skype, online video conferencing, real time chat systems, social networking sites such as Facebook.

Automatic information-harvesting tools: include news (RSS) feeds, social tagging and bookmarking tools, social networking sites that gather and compare information about users and their patterns of behaviour, institutional repositories that harvest documents from desk tops.

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.

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.

ICT: information communication technologies, typically, although not necessarily Internet-based use of technology to process information and aid communications.

Portal: a portal places the emphasis on federating distributed sites of information using technologies such as cross searching, harvesting, and alerting, and collates this into an amalgamated form for presentation to the user.

Resource discovery tools: such as search engines, subject gateways, RSS feeds, automated data aggregators, ranking systems, help users to find relevant information more efficiently by searching multiple databases, filtering results and providing alerts regarding information updates.

Workflow management tools: the idea of 'workflow' encompasses a whole cycle of activity, not just a particular task, and may extend across the responsibilities of several people, Thus a 'digitization workflow' may include a photographer, archivist, curator and encompass acquisition and cataloging of materials, preparation for digitization, equipment calibration and operation, file naming and type selection, storage type selection, metadata schema application and physical handling, storage and return of the artefacts. Workflow management tools such as spreadsheets, project management schedulers, content management systems, version control systems, tracking systems, bookmarking desktop indexing and searching systems and annotation tools help users to organize the workflow, to optimize it and to ensure quality standards and procedures are met.

References

Brewer, J. and Kilbride, W. G. (2005). *HEIRNET User Survey 2005 Report and analysis*. York, UK: Council for British Archaeology.

Brown, S., Ross, R., Gerrard, D., Greegrass, M., Bryson, J. (2006) *RePAH: A User Requirements Analysis for Portals in the Arts and Humanities Final Report*. Humanities Research Institute Online, Sheffield University, UK ISBN 0-9542608-8-0. http://repah.dmu.ac.uk/report/index.html (accessed 10 June 2009).

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/summarys/html98_9/2178main.pdf (accessed 10 December 2008).

Cultural Heritage Consortium (2002). *Heirnet: Historic Environment Information Resources Network. Users and their Uses of Heirs*. http://www.britarch.ac.uk/HEIRNET/publications.html (accessed 10 December 2008).

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/(accessed 10 December 2008).

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 (accessed 10 December 2008).

Greengrass, M. (2006). *RePAH Research Portals in the Humanities: a user analysis project. Appendix A3: The Development of Portal Provision in the Arts and Humanities, 1996–2006*. Sheffield: HriOnline http://repah.dmu.ac.uk/report (accessed 10 December 2008).

Greenstein, D. (1998). The Arts and Humanities Data Service Three Years' On', *D-Lib Magazine*, **12.** http://www.dlib.org/dlib/december98/greenstein/12greenstein.html (accessed 10 December 2008).

Grout, C. and Rymer, J. (1998). *VADS User Needs Survey 1998: Report*. http://vads.ahds.ac.uk/reports/user_survey/user_survey.html (accessed 10 December 2008).

Guy, M. (2003). *User Testing Report. Subject Access to the DNER (SAD I)*. http://www.portal.ac.uk/spp/documents/testing/phase1/usertestingreportv3.doc (accessed 10 December 2008).

Hiom, D. (2006). Retrospective on the RDN. *Ariadne*, **47.** http://www.ariadne.ac.uk/issue47/hiom/intro.html (accessed 10 December 2008).

Hix, D. and Hartson, H. R. (1993). *Developing User Interfaces: Ensuring Usability Through Product & Process*. New York: Wiley.

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/ (accessed 10 December 2008).

Rettig, M. (1994). Prototyping for tiny fingers. *Communications of the ACM*, **37**(4): 21–27.

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 (accessed 10 December 2008).

Stone, S. (ed.) (1982). Humanities Information Research: Proceedings of a Seminar; Sheffield 1980. Centre for Research on User Studies, *CRUS Occasional Paper* 4.

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/bm~doc/unsworth.pdf (accessed 20 June 2008).

Warwick, C., Terras, M., Huntington, P., and Pappa, N. (2008). 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. *Literary and Linguistic Computing*, **23**(1): 85–102.

Weaver, W. T. (1971). The Delphi Forecasting Method. *Phi Delta Kappan*, **52**(5): 267–73.

Young, R. (2004). *Findings of the Altis Survey 2003–2004*. http://www.portal.ac.uk/spp/documents/testing/phase2/altis/AltisSurvey20032004.doc (accessed 20 June 2008).

Notes

1 Intute web site http://www.intute.ac.uk/about.html
2 http://www.ariadne.ac.uk/issue47/hiom/intro.html#introduction
3 http://ahds.ac.uk/about/ahds-timeline.htm
4 (http://www.jisc-collections.ac.uk/)
5 http://www.methodsnetwork.ac.uk/
6 http://www.ahessc.ac.uk/ahessc-home
7 http://www.arts-humanities.net/
8 http://www.ahrcict.rdg.ac.uk/new_projects/networkofcentres/index.htm
9 http://www.jisc.ac.uk/whatwedo/programmes/portals/faq.aspx