



# RePAH:

## Research Portals in the Arts and Humanities

A user analysis project

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

Report prepared by Mark Greengrass

<http://repah.dmu.ac.uk/report>



Arts & Humanities  
Research Council

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

Published by:

**HriOnline**

**for**

**The RePAH Project**

Knowledge Media Design	&	The Humanities Research Institute
De Montfort University		University of Sheffield
Portland 2.3a		34 Gell St
The Gateway		Sheffield South Yorkshire S10 2TN
Leicester LE1 9BH		

ISBN: 0-9542608-8-0

**Also Available at** <http://repah.dmu.ac.uk/report>

The right of Stephen Brown, Robb Ross, David Gerrard, Mark Greengrass and Jared Bryson to be identified as the Authors of this Work has been asserted by them in accordance with the Copyright, Designs and Patents Act of 1988.

September 2006



## Contents

Contents .....	3
3.1 The Concept of e-infrastructure .....	6
3.2 The Evolution of RDN Subject Portals.....	7
3.3 User-Requirements Analysis and RDN Arts and Humanities Developments .....	11
3.4 The Evolution of the AHDS .....	13
3.5 Other Relevant User-Needs Requirements Analyses .....	17
3.6 RDN and AHDS Financial Support.....	18
3.7 RDN and AHDS: Complementary Resource Discovery Agents.....	19

## **The RePAH 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 sought to understand how the arts and humanities research community finds and exploits the internet resources it needs.

In order to do this the RePAH project:

- Examined the existing literature on user needs with regard to web gateways and portals,
- Analysed the web-logs from the Arts and Humanities Data Service (AHDS) subject centres and the Resource Discovery Network's (RDN) humanities and arts web hubs (prior to July 2006 these were known as Humbul and Artifact, but now have been harmonised into Intute-Arts and Humanities)
- Conducted focus groups, interviews and a Delphi exercise with members of the arts and humanities community
- Developed and tested a paper-based demonstrator for a managed research environment to explore possible ways forward with regard to web-based research resources.

The project was carried out in 7 work packages:

- WP1 RePAH Online Questionnaire--this report examines an online survey of the Arts and Humanities Community's use of web resources.
- WP2 Web-Log Analysis--this report analyses web-logs from several of the Arts and Humanities Data Service subject centres as well as Humbul and Artifact of the Resource Discovery Network (now Intute).
- WP3 First Focus Group--this report studies the responses from a series of five focus groups conducted at the University of Sheffield and three interviews from DeMontfort University. Respondents discussed their use of web resources in general and portals in particular.
- WP4 Delphi Exercise--this report considers the results of a Delphi exercise conducted around the feasibility of various web-based tools.
- WP5 Demonstrator of a Managed Research Environment--this report is an exploration of a paper-based demonstrator of a variety of features that might be applied as portlets and used by the Arts and Humanities research community.
- WP6 Phase II User Trials of Portal Demonstrator--this report brought the paper-based demonstrator to scholars in eight subjects within the Arts and Humanities community and asked them to evaluate the features and functionality of possible portlet tools.
- WP7 Intute in Light of this Report--this report explores Intute-Arts and Humanities with reference to the features and functionality explored in the paper-based managed research environment demonstrator.

Additional appendices within the RePAH Project report include an overview of the Arts and Humanities research community [Appendix A2], and a review of the literature relevant to user requirements for digital resources and web-based research facilities [Appendix A3].

This appendix reports on Work Package 7 which examines Intute-Arts and Humanities with reference to the features and functionality explored in the paper-based managed research environment demonstrator, as well as some the data harvesting of the AHDS by Intute.

To see the full report and the other appendices see <http://repah.dmu.ac.uk/report>.



### 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:

- a **baseware** ‘layer of **base technologies**...the integrated electro-optical components of computation storage, and communication’
- a **middleware** ‘layer of **enabling hardware**, algorithms, software, communications, institutions, and personnel’ that lie between’
- 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:

- the need to create ‘access’ and ‘discovery’ services.
- the belief in centralised provision of such services as an emerging ‘academic infrastructure’ within a relatively coherent UK higher educational framework.
- 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.
- 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, teaching 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:

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

- 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:

- Provision of RSS news feeds relating to Humbul database content
- Provision of third-party RSS feeds (but only jobs.ac.uk provided)
- Email alerting service
- User profiling and improved saved search functionality
- 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:

- Cross-searching of remote arts and humanities databases
- 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:

- Harvesting and indexing of third-party OAI metadata
- Provision of Z39.50 service (provided for a time, but then taken out of operation)
- Provision of public OAI metadata repository (provided for a time, but taken out of service)
- Provision of an Events database (proposed for inclusion but incomplete development)
- 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:

- What do you like about the portal?
- What don't you like?
- What should work better?
- Would you use the portal for your own research?
- 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:

- The interdisciplinary component of Humbul was appreciated. Cross-searching 'could be very popular' but 'whether I would ever use that I do not know'.
- 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'
- 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
- 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:

- I mainly use Google search the web.
- It would be good to see a variety of news sources in one location.
- I would like to see a conference and events listing.
- Email alerts of new resources would be something I would like to see.
- I would use a service that searches multiple databases from one location.
- Personalisation of a website is something I would use if it had the right options.
- A full text journal search is something I would use.
- 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:

- Preserving arts and humanities digital resources created by Higher Education
- **Providing rich, deep, access to the intellectual content of arts and humanities digital resources created by and for Higher Education**
- 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]:

- Archaeology Data Service – now AHDS Archaeology [York]
- History Data Service – now AHDS History [Essex]
- Oxford Text Archive – now AHDS Literature, Language and Linguistics [Oxford]
- Performing Arts Data Service – now AHDS Performing Arts [Glasgow]
- 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](http://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:

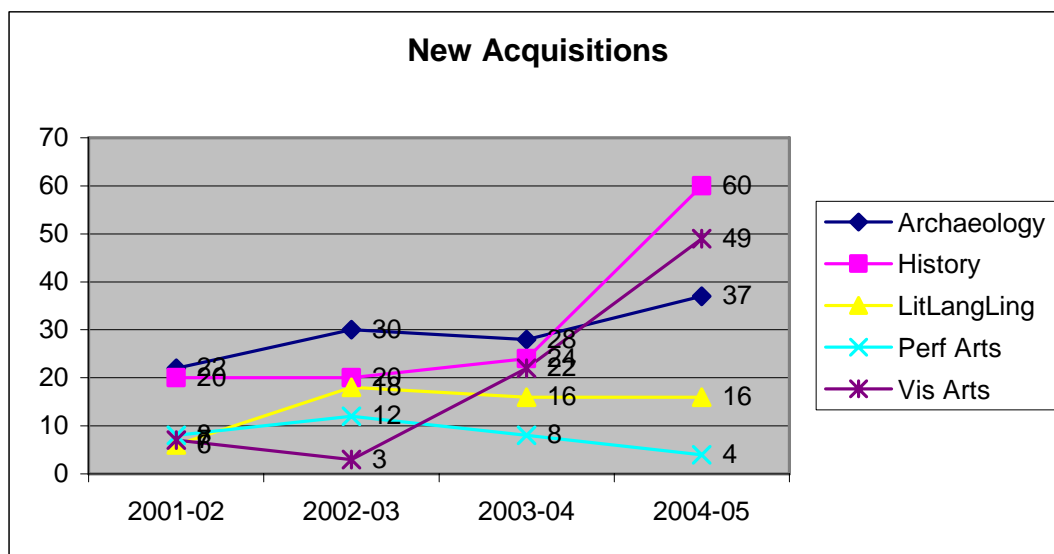
**Figure 1**

**Total number of new acquisitions**

2001-02	<b>63</b>
2002-03	<b>83</b>
2003-04	<b>98</b>
2004-05	<b>166</b>



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:

- A user-needs survey was conducted by the Visual Arts DS in December 1997-February 1998, based on a paper and online questionnaire, to which it had 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

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 [Condrón, 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:

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