Treatments for *'Agoraphobia'*: International Developments for the Hybrid Library

Vincent Galante
Head, Information Systems
La Trobe University Library
V.Galante@latrobe.edu.au

Abstract:

This paper based on a study visit made in May 1999, generously sponsored by the VALA Travel Scholarship program, reports on developmental projects that seek to aid library users to efficiently discover and locate information resources. In the current information landscape, where resources exist in a variety of formats, the goal is to develop systems that seamlessly integrate the globally distributed content. The report includes descriptions of the most promising UK eLib program funded research projects into Hybrid Library Systems and some of the Digital Library Initiatives in USA Universities. It also identifies and provides links to other international research on creating Digital Library Systems.

Introduction

Agora, n. the public square and market place in an ancient Greek city.

Agoraphobia - fear of public places; fear of open spaces; fear of virtual spaces

Background to Agoraphobia

This paper goes back to late winter in 1998 at La Trobe University. The Library's entrance, or exit depending on your perspective, is onto the "Agora" - our very own public square where I would briefly escape during the more stressful periods of the implementation of our new Integrated Library Management System (ILMS). As project manager for the implementation I was sometimes stricken by a phobic reaction, let us name it "agoraphobia" for the place where it was felt and for the sake of this story. When I finally realized that the implementation was progressing smoothly, on schedule and within budget, my breaks became islands of contemplation. It was at this time that I saw the advertisement for VALA's Travel Scholarship and an opportunity to explore the new frontiers. And it was during one of these breaks that I realized: the new ILMS of which we were so proud, was a long way from "integrated" to the Library's users.

The unfortunate Information Seeker has to understand when to search the OPAC, the Coolcat, other library catalogues, the numerous standalone or networked CD-ROM indexing and abstracting databases, web accessible databases, full text databases, datasets, locally mounted web resources, remote internet resources – you get the picture! I imagined a poor undergraduate stumbling his or her way into their subject appropriate indexing and abstracting database and finding the perfect reference only to be confronted with the following dilemmas:

"Having found a relevant reference how do I get to see it? Oh! I need to search the catalogue to find the journal title. But the journal title isn't in this catalogue. Ah! I need to search the full text database or other library catalogues. No luck. Hmm! Maybe that journal is directly published electronically and can only be accessed via the electronic journals webpage. But it's not there. Okay, I'll scribble the reference down and go and fill out an Inter-Library-Document-Delivery form. Damn! I'm not eligible for an ILDD. So I'll search the Internet. Groan! Looks like that search engine wasn't comprehensive. I guess I'll have to wait for Godot - or is that Google?"

Very few of our users are technophobes, and it is not information they fear. It is more an anxiety over how to navigate this virtual information space that they can't see and can't map. And then there's the language - every second word is new or an acronym (sorry, but just look at this paper). No wonder many of our students think, "It's all Greek to me and this information market place is just too vast. Take me away, I'm Agoraphobic!"

In developing my application for a VALA Travel Scholarship, I needed to understand what modern libraries were. Computing machines have been transformed into Information Technologies but what have libraries been transformed into? The occasional one has been renamed an Information Centre but generally the word "library" persists even for entities that are fully digital. Therefore, we have terms like Digital Library, Electronic Library, Virtual Library, Hybrid Library and even Cybrary. Personally, I prefer the term Hybrid Library because it reflects the many forms in which information can exist. It also suggests the evolving nature of

information resources, from physical artifacts in physical buildings to globally accessible digital forms.

My Travel Scholarship proposal was to survey developments that integrated information resources and to explore technologies that assisted library users to more easily find and locate their information needs. Using the reference tool of choice - The Web - I fortuitously discovered two projects that appeared to be spot on cures for "agoraphobia" and hundreds of other projects that might have some curative powers.

The two prime candidates for visits were:
      "Agora - An eLib Hybrid Libraries Project" [1]
      "Agora - Your all-in-one digital Library" [2]

Other projects visited included:
      California Digital Library – CDL [3]
      Stanford Digital Library Initiative – Infobus [4]
      Berkeley Digital Library SunSITE [5]
      Digital Library Federation Project- Making Of America II [6]
      Berkeley University, Finding Aid Project [7]
      Berkeley University, Chesire II Search Service [8]
      New York University Medical Library - Ehrman Digital Library [9]
      NYPL: The Science, Industry and Business Library Electronic Information Centre [10]
      eLib Projects – Models, Headline, BUILDER, M25, and ROADS [11]

A short conference paper cannot do justice to any of these projects therefore I will provide you with snippets from some of my favourite projects and most pertinent to my quest. Hopefully you will be enticed to visit these and the other sites virtually, if not in body as I did. Most of the projects visited have well developed web sites, even those in the early phases of project developments.

The eLib Projects

In 1995 the Joint Information Systems Committee (JISC) in the UK established the Electronic Libraries (eLib) Programme [11] to help develop and implement technologies in the Higher Education community to better integrate traditional and non-traditional (electronic) resources within the library environment. Phase 1 and 2 of eLib funded about sixty projects to the tune of £15 million over three years. In 1998, a further £10 million was made available for Phase 3 Projects. These were to consolidate the knowledge gained from the previous Phases and to build the digital libraries of the future.

The 12 new and 15 old projects in Phase 3 covered the following broad areas:
      Hybrid Library Systems
      Large scale resource discovery (Clumps)
      Subject-Based Internet Gateways
      Electronic Document Delivery
      Electronic Journals
      Digital Preservation

While all the projects seemed of great interest and potential I had limited time and needed to focus on a few. With the help of discussions with Chris Rusbridge, eLib Programme Director, I chose to focus on three Hybrid Library Systems (Agora, BUILDER and Headline), one Clumps project (M25), and one Subject-Based Internet Gateways project (ROADS).

AGORA

The AGORA project [1] has the aim of seamlessly integrating a variety of information sources in a user-centred search and retrieval system, with the ultimate goal of developing a standards-based Hybrid Library Management System (HLMS). This is a collaborative project with the University of East Anglia, the Centre for Research in Library and Information Management (CERLIM) at Manchester University, Fretwell-Downing Informatics Limited (FDI) and is centred at the UK Office for Library and Information Networking (UKOLN), a research centre at Bath University.

AGORA is an attempt to put into practice many of the concepts developed by another eLib UKOLN project - MODELS. The main outcome from MODELS (Moving to Distributed Environments for Library Services) was the development of a model systems framework to manage heterogenous services. This has come to be known as the MODELS Information Architecture (MIA). MIA provides a common vocabulary for both systems developers and information managers with the central concept being the broker (market place) model for the exchange of standardized data.

At the time of my visit (May 1999) the project was only approaching the end of year one of its three year life. Much of the work I was able to assess was still at a theoretical stage. However, work had commenced with their commercial partner to enhance the FDI Virtual Document Exchange (VDX) system using the MIA framework for a demonstrator system Release I in September 1999 and Release II in August 2000.

The AGORA system is based on a four-layer model as in the following diagram.
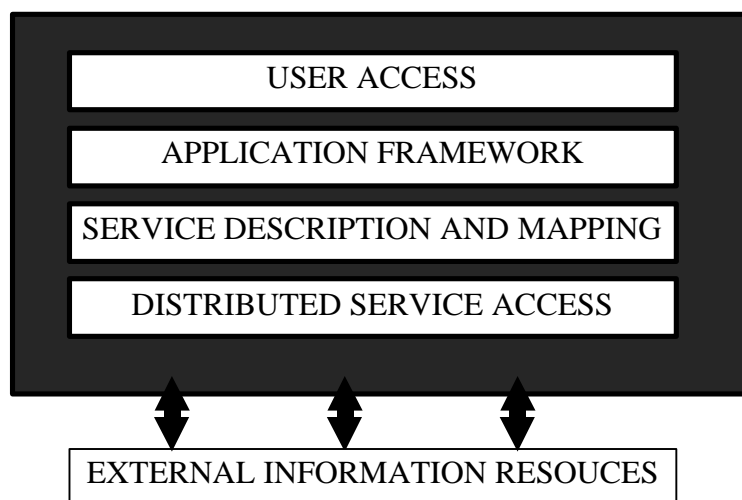
USER ACCESS - manages the interaction with the user by providing a coherent "Information Landscape" which is dependent on the user profile.

APPLICATION FRAMEWORK - provides the "discover", "locate", "request" and "deliver" functions.

Discovery may be a complex and iterative process involving many levels and is dependent on layer 3 providing a searchable database of databases. One first needs to discover the applicable resource services and then the information object.

Location is the identification of the required information object. This could be a holdings statement from a library catalogue or some persistent identifier for an electronic object.

Request uses the data from the location function (ie. no need to rekey details) to manage the request/order function. This could be as simple as instructing the user to go to the library shelves or executing a HTTP "get" command but more typically it involves passing the transaction onto an ILL or commercial supplier using the evolving standards based protocols.

Delivery is the transfer of the item to the user hopefully over HTTP but it could be a physical item from a local collection, a postal delivery or a faxed item.

While all these functions are typical of most systems the aim of AGORA is to present an "organised space" to the information seeker, hide the underlying differences between systems and seamlessly pass data from one system to another.

SERVICE DESCRIPTION AND MAPPING - This layer holds information about the services that have been configured for this particular "Information market place". In particular it holds the collection level descriptions. The collections might be library catalogues, indexing or full-text databases, subject gateways, web sites or document supply centres. How to characterise collections is a poorly defined area at the moment and a standards approach is urgently needed. Good collection descriptions provide the user with knowledge about the available resources but more importantly allow for effective cross-domain searching. The aim is to pinpoint only the most relevant services and thereby avoid a brute force approach that will both overload the distributed systems and the user with many irrelevant items. This layer is also responsible for maintaining the application interface description for each service. These descriptions allow the AGORA system to machine interface with the various distributed propriety systems.

DISTRIBUTED SERVICE ACCESS - This is the software layer that does the actual search, request and deliver functions on the remote heterogeneous systems. AGORA will only focus on Z39.50 and Whois++ searchable systems and request/delivery systems using the ISO-ILL, Z39.50 item order, or the HTTP "get" protocol. While the AGORA team is reluctant to interface with non-standards based systems, the model can be adopted to interface with propriety databases of which there are many, especially in the CD-ROM database environment.

In summary, the AGORA project is an attempt to build a broker based Hybrid Library Management System that provides for:
- a "plug and play" environment for libraries allowing them to connect their own selection of heterogeneous, distributed, and ever increasing information resources
- a standard user interface linking the heterogeneous resources in an information landscape dependant on individual user profiles
- cross domain searching which hides the differences and the mechanics of interaction from users
- collation of search results from these varying systems
- flow of data from one system to another so as to automate the locate, request and delivery processes.

I strongly recommend that you follow this project on their web site; because it is standards-based it has the potential to live up to early expectations. While you are at the AGROA web site I would also suggest looking at some of the other UKOLN projects. During my week's visit there I found the Research Centre an intellectually stimulating place and a hot-bed of development with many projects cross-fertilizing each other. The following lists some of the other projects that I encountered at UKOLN, though not at the same level of detail as my prime interest AGORA.

ROADS - Resource Organisation and Discovery in Subject-based services. This is another eLib project to design and implement a standards-based subject gateway to Internet resources. In particular look at their review of products designed to enable resource discovery. <http://www.ukoln.ac.uk/roads/product-comparison/> You can also get a copy of their open-source UNIX software resource discovery tools from the site.

CEDARS [12] - The CURL Exemplar for Digital ARchiveS is an eLib project aiming to address strategic, methodological and practical issues relating to the long-term preservation of digital information. <http://www.ukoln.ac.uk/CEDARS/>

BIBLINK [13] - A project to enable selected electronic publishers to transmit metadata to national bibliographic agencies and to convert this data into MARC.
 <http://www.ukoln.ac.uk/metadata/BIBLINK/>

PRIDE [14] - People and Resource Identification for a Distributed Environment.
<http://www.ukoln.ac.uk/metadata/PRIDE/>

The Web Watch Project [15] surveys emerging web technologies and is part of the Web-Focus Group at UKOLN. <http://www.ukoln.ac.uk/web-focus/webwatch/>

HEADLINE

HEADLINE [16] is another eLib project in the Hybrid Library System stream and is based at the London School of Economics. While it shares many of the goals of the AGORA project, it was at an earlier phase of development when I visited the project team and had more modest aims. HEADLINE aims at developing a working model for consistent access to library materials regardless of physical form. One of its stated aims is to produce a "Toolkit of Techniques" which accommodates print, electronic and other sources and focuses on the needs of different user groups so as to simplify the increasingly confusing choice of information sources.

A key and distinguishing feature of the HEADLINE Hybrid Library System is development of the Personal Information Environment (PIE). This is the interface where a user's interests and rights are matched against the subject content and licenses of available resources. This narrows the large set of resources to search and reduces the large and often irrelevant or inaccessible result set. To achieve this the system maintains two databases using generic SQL techniques, thus allowing for transportability to any future non-propriety system. The system for matching user rights against resource grants is an integral component of the hybrid library gateway. [Libraries have complex sets of rules for levels of access to their resources - just look at any library's loan rules.] The User Data Module (UDM) maintains the users' administration status, subject areas, courses, personal interests. The Resource Data Module (RDM) maintains subject

metadata and licence agreements for that library's set of resources. In order not to re-invent the wheel, HEADLINE's project manager is hoping to utilize existing technologies particularly those developed for e-commerce such as portals like My.Yahoo!

The only demonstration at HEADLINE that was working when I visited was the Shelfmark & Resource Locator (SHERLOC). This is a resource 'discovery-to-access' tool developed in an effort to 'hybridise' the library print collection by making location of shelved items easier. This was developed for LSE because of the large number of "Where is ...?" questions at the library information desks. Currently it requires the user to input a Classmark into a web form and a plan of the appropriate floor of the library is returned to their screen, with an 'X marks the spot' showing the shelf location. In the future, it is hoped to create a 'live link' between the library catalogue and the Hybrid Library System Interface.
See: <http://decomate.lse.ac.uk/headline/demo.html>

BUILDER

BUILDER [17], the Birmingham University Integrated Library Development and Electronic Resource, was the third eLib Hybrid Library System project I visited. Their stated aim is to develop a "working model" of the hybrid library. BUILDER is different from the other projects in that it is very much an institutional approach. They only wish to connect their own services together as they do not want to "pull bits from everywhere" or from systems they have no control over. They have adopted a very pragmatic approach and will use standards-based approaches only where they currently exist and will write their own interfaces where none exist in order to connect their existing systems.

BUILDER is developing six prototype modules:
- user registration and induction
- ordering and delivery of materials
- metadata index to printed and electronic resources
- teaching and learning resource packs
- publication and digitisation
- access gateways

As with the other eLib projects I visited, developments were still at an early stage but it will be an interesting project to follow as they had very clear objectives of how to integrate their own resources. They also have project team members with excellent technical skills and good access to other IT specialists. Birmingham University has merged their Library and Computer Centre. The integration is reported to have been highly successful partly because it has happened at all levels of the hierarchies and because Librarians and IT personnel work side by side. The new entity has a strong commitment to implementing successful technologies and this should bode well for BUILDER.

Two technical developments I did see were an API toolkit which provides hooks into their Library Management System - Talis, and secondly, a Java thin client to their proprietary CD-ROM databases. Their CD-ROM network, ATLAS, runs under Windows NT using the Citrix Winframe server and their achievement has been to create a Java API to link into the CD-ROM network. This allows any user on any platform to access the databases using a Java enabled Browser. The BUILDER demo web site at <http://builder.bham.ac.uk/demo>. allows you to trial this technology.

M25

M25 [18] is an eLib Clumps project. 'Clumps' are virtual union catalogues and the M25 is the motorway around the Greater London area. The aim of the project is to allow users to search across all the 39-member M25 Consortium Libraries and to be able to limit searches either by geographic area or by subject strength of the library. The subject list consists of only 125 terms but it will be enhanced to become a metadata collection level search tool. The M25 Clumps interface will also be enhanced to return user access policies and opening hours for the searched libraries.

The project has successfully searched across the six pilot libraries and covered searching the five most prominent library management systems: SIRSI, INNOPAC, TALIS, HORIZON and LIBERTAS. However early testing seems to indicate that it might not scale well due to higher probabilities of finding at least one slow network connection and problems de-duplicating large result sets. While other eLib clump projects are using commercial Z39.50 clients, in particular the Fretwell-Downing VDX client, the M25 project team have enhanced the open source free Z39.50 client - Europgate and have had great success with it. The project manager indicated that they are willing to share their enhancements with any site wanting a copy.

California Digital Library

The CDL [3] was not on my itinerary when I applied for the VALA travel scholarship in late 1998 as it only opened its virtual doors on January 30, 1999. However it was one of the most functional and well thought out sites I visited. CDL is described as the tenth library of the University of California System that includes notables such as Berkeley University, UCLA, and UCSD. It is a serious attempt with serious money at providing a working digital library that only takes on services already proven in some other test-bed. Its primary purpose is to design and create systems that support the shared digital collections of the University of California. CDL also encompasses all the activities of the former Division of Library Automation, which includes the maintenance of the Melvyl Catalogue and locally hosted commercially-produced databases such as MEDLINE and INSPEC.

In this short conference paper I cannot do justice to all the services provided so I recommend a visit to their web site <http://www.cdlib.org/about/> to discover more. However, I would like to mention three aspects of the site which I found innovative and which make it rate as a good treatment for agoraphobia. Firstly, access to resources and customization of the user interface are dependent on user profiles. Your user profile also provides the value-added feature of allowing you to save your searches and having the SDI (Selective Dissemination of Information) results emailed to you at regular intervals. Secondly the collection of resources at CDL is searchable, (it's too big a site to navigate). This meta-database is being developed using evolving metadata standards for collection level descriptors.

Finally, the best thing at CDL is the ability to search the indexing and abstracting databases and to not only see which UC libraries hold the print journal but to be able to move from the database to the electronic version of the article in a single click. This ability to click through to the article level was first made available on the CDL locally mounted databases but they were planning to extend this feature to databases located at vendor sites as well. Although the magic was not fully explained to me, it works in the background by a Z39.50 lookup to a CDL meta-database of more than 2400 electronically published journals and returns on the fly the article level URL. More information can be found at <http://www.cdlib.org/>

Digital Library Initiatives

The Digital Library Joint Research Initiative, through the NASA, NSF and ARPA, jointly funded six research and development projects. The funding primarily went to computer science departments to address the digital library life cycle from information creation, access and use, to archiving and preservation. The two projects I visited were the Stanford Infobus project and the Berkeley Chesire II project as both were working with interoperability issues.

InfoBus

InfoBus [4] the Stanford DLI project had a different mission to the UK eLib projects. It was more research led and with a strong computer science focus. In fact, the grant was used primarily to fund a significant number of PHD theses. The focus was also more 'digital' than 'hybrid'. While the project's objective was to enhance accessibility among distributed information repositories the InfoBus project had a very explicit systems architectural focus. The InfoBus infrastructure was designed to provide interoperability among heterogeneous, autonomous software components, much like the computer hardware bus enables interactions between the different hardware components on a computer motherboard. In particular the project pursued a suite of related investigations within a distributed object environment, based on the CORBA (Common Object Request Broker Architecture) standard. CORBA is a mature standard introduced in 1991 by the Object Management Group (OMG) and defines the Interface Definition Language (IDL) and the Application Programming Interfaces (API) that enable client/server object interactions.

While much of the work was highly technical and beyond my level of comprehension, a number of sub-projects were of particular relevance. They succeeded in building a test interface which permitted users to interact with many information sources (The Dialog service, MIT press, the ACM, the Web and Stanford's Library Catalogue) without having to be aware of the diverse query formats. This system also allowed many users to interact with and annotate documents while preserving the original. They are also developing an information filtering service prototype known as the Glossary of Servers Server (GLOSS), which assembles metadata from disparate resources so that user queries can be more directed. Another sub-project is the Stanford Information Filtering Tool (SIFT) which matches user-interest profiles against large numbers of documents.

Stanford University has also secured funding from the Digital Libraries Initiative II and will continue the research of the most promising projects. Additionally, with this round of funding they are developing the Simple Digital Library Interoperability Protocol (SDLIP-Core), the outcome of which might revolutionize searching across distributed systems. To follow the progress of this protocol see <hhtp://DLIB2.STANFORD.EDU/~testbed/doc2/SDLIP>

Cheshire II Search Service

Cheshire II is an experimental online catalogue system comprised of a Z39.50 client and server. It was designed to extend the realm of existing online catalogue systems, which only search bibliographic (MARC) databases, and to encompass the full range of full-text and multimedia online resources. The main objectives have been: to design a user interface with advanced search facilities in a library web environment at the client end; and to develop a Z39.50 server that understands database records that are non-MARC and that can easily adapt to new types

of data. In the design of Cheshire II the project team endeavoured to use only recognised national and international standards. For the database they chose to use the Standard Generalized Markup Language (SGML) as the data storage type and adopted the SGML Data Type Definition (DTD) language to define the structure of the data. Therefore all the data in a Cheshire II server is stored as tagged SGML documents. This includes MARC records, for which they have developed a MARC to SGML converter.

Although Cheshire II is still a Computer Science Laboratory prototype, I saw that it had much potential, particularly considering that much of the publishing industry routinely uses the SGML standard. This should enable libraries to acquire from numerous sources SGML tagged digital objects and load them into their own servers, thereby preserving the custodian role of libraries.

Digital Library SunSITE

While at Berkeley University I also visited with the manager and staff working on the Berkeley Digital Library SunSITE [5]. <http://sunsite.berkeley.edu/> In the continuum of library - hybrid library - digital library, SunSITE sits very much at the digital end. No library catalogues to search, no indexing and abstracting databases, no distributed electronic collections - just a treasure trove of electronic resources, from discussion lists to full-text books. [See The Jack London Collection <http://sunsite.berkeley.edu/London/>]

From the perspective of my travel goals, it was nevertheless an exemplary site to visit as they are working with and developing a variety of resources to determine the best practices for creating digital collections and services. Two things of particular note are the SWISH-Enhanced indexing/search software and the Digital Library Information Resources section.

SWISH-E (Simple WAIS Indexing System for Humans – Enhanced) <http://sunsite.berkeley.edu/SWISH-E/> is a powerful freely available indexing and search tool with full support for HTML META tags. This tool, in combination with a metadata standard such as Dublin Core, can assist digital librarians to build catalogues of digital materials searchable either as full text or via fielded searches.

The Digital Library Information Resources section <http://sunsite.berkeley.edu/Info/> is the most comprehensive site that I have encountered for any budding digital librarian. SunSITE's mission is two-fold, to create digital library collections, and equally importantly, to help others do the same. SunSITE's goal is to locate the best information on building digital libraries and to either host it locally or to point to it. (It was my first port of call when I was planning for the VALA Travel Scholarship.)

AGORA - Your all-in-one Digital Library Solution

I selected this site not only because of its name, but because it gave me an opportunity to explore some of the developments in mainland Europe that rarely get reported in mainstream English publications. The Digital Library System Agora [2] was developed by the commercial company Satz-Rechen-Zentrum (SRZ) in Berlin, Germany in collaboration with the Goettingen University Library. Essentially, Agora is a Document Management System (DMS) and not directly related to the core theme of this paper. However, because it has many impressive features, is very much standards-based, and has a reasonable price tag, it deserves some

mention particularly for those wishing to create a "digital library". The administrative tools provide the following features:

- importing/exporting of digital documents into any relational database (scanned images and metadata in XML/RDF and PDF formats)
- describing the document structure tree and the image set
- mapping bibliographic and structural data to a set of images
- batch conversion of TIFF images at the point of input or on-the-fly for web access
- generation of WWW pages from HTML templates for objects like table of contents or image views
- simple or advanced fulltext searching using the Verity search engine.

For a more in depth exploration of Agora's capabilities see the web site. SRZ have promised that more of the details will become available in English. <http://www.agora.de/>

CONCLUSION

Information resources now available in print, electronically or through global networks are anything but organised. No matter how information is published and collected, if it is to be made readily available to our users, there must be general agreement about the requirements for systems architecture, metadata, indexing and retrieval standards. I am excited and positive about many of the projects I saw. However, it will require significant additional effort and exploration to develop and adopt common standards in order to achieve interoperability between systems, let alone to provide our library users with quick and direct access to the most relevant and useable information resources.

Acknowledgments

References

[1] Agora - An eLib Hybrid Libraries Project,
http://hosted.ukoln.ac.uk/agora

[2] Agora - Your all-in-one digital Library
http://agora.com.de

[3] California Digital Library - CDL
http://www.cdlib.org

[4] Stanford Digital Library Initiative - Infobus
http://www-diglib.stanford.edu/

[5] Berkeley Digital Library SunSITE
http://sunsite.berkely.edu

[6] Digital Library Federation Project- Making Of America II
http://sunsite.lib.berkeley.edu/MOA2/

[7] Berkeley University, Finding Aid Project
http://sunsite2.berkelyedu/oac/

[8] Berkeley University, Chesire II Search Service
http://128.32.224.173/chesire/

[9] New York University Medical Library - Ehrman Digital Library
http://www.nyu.medlib.edu/

[10] NYPL – The Science, Industry and Business Library Electronic Information Centre
http://www.nypl.org/SIBL/

[11] eLib Projects – including Models, Headline, BUILDER, M25, and ROADS
http://www.ukoln.ac.uk/services/elib

[12] CEDARS [12] - The CURL Exemplar for Digital ARchiveS
http://www.ukoln.ac.uk/CEDARS/>

[13] BIBLINK
http://www.ukoln.ac.uk/metadata/BIBLINK/

[14] PRIDE - People and Resource Identification for a Distributed Environment.
http://www.ukoln.ac.uk/metadata/PRIDE/

[15] The Web Watch Project
http://www.ukoln.ac.uk/web-focus/webwatch/

[16] HEADLINE – Hybrid Library System
http://decomate.lse.ac.uk/headline/

[17] BUILDER - Hybrid Library System
http://builder.bham.ac.uk/

[18] M25 – An eLib Clumps Project
http://www.M25lib.ac.uk/M25/