

Not just a portal: Managing Access in a Complex Information Environment

Jean Sykes, John Paschoud, Christine Cooper

London School of Economics & Political Science, United Kingdom
{j.sykes , j.paschoud , christine.cooper}@lse.ac.uk

Abstract

Work is in progress by LSE staff to create a Managed Information and Knowledge Environment, simple and managed access to a wide range of appropriate and permitted content for a broad range of users. The portal is only one small part. More difficult tasks are identifying information content (internal and external) for different user types, and developing suitable middleware for managing access, in an institutional, national and international context. Two projects will be highlighted: SECURE, working on the access management middleware, and UK Computing Plus, offering electronic information access to certain library visitors.

Keywords: access management; integration; authorisation.

1 Introduction

Everyone is talking about portals. Software and systems suppliers are creating portals as part of their product portfolios, and universities across the world are becoming aware of the need to have an institutional portal.

Why the need? Because researchers, students and administrators are using more and more electronic access to content across a wide spectrum of information types:

- institutional and third party
- commercial and open source
- stored locally and stored remotely (anywhere in the world)
- quality assured and not
- primary and secondary
- in digest form and full text

Most of this content is available via the web, but there are also many information sources – especially those stored within an institution for internal purposes only – which are still paper-based, or hidden in personal/group databases, or filed in such applications as Microsoft Exchange® public folders. Increasingly users want all the information relevant to their work to be available to them electronically on a 24 x 7 basis and from wherever they happen to be. They do not want to have to log in and out of different systems with different logon procedures and passwords; they need to navigate between their sources. They also want to eliminate

duplication of effort (for example re-keying the same search in different databases) and they want irrelevant information to be filtered out for them. But what is irrelevant to a user today may become relevant for them tomorrow, and so they want flexibility and inclusivity as well as refinement and exclusivity. This is a tall order for information providers to fulfil and is a major challenge for us all.

2 The vision for LSE

At the London School of Economics and Political Science (LSE) we are keen to provide our users with a portal to support their specific needs. This means creating different “views” to different users, and it means allowing users to access discrete parts of the information environment if they wish to as well, since always entering through a particular view may not be the quickest and most efficient way in for certain tasks. Unlike many of our colleagues in other universities, therefore, we do not entirely subscribe to the single portal vision.

Moreover, our definition of a portal is that it is just a gateway to the information, just a part of the whole which we are trying to create. The portal is the front-end, the user interface, and of course we do need a new generation of such applications to cope with the complexity of the new need for several user views. But ultimately the portal is the easy bit. Far more difficult are the other two major areas involved in building this information environment for our users:

- the information content at the back end
- the middleware which facilitates dialogue between the portal at the front and the information content at the back

So at the LSE we are not calling our project a portal. We are calling it a Managed Information and Knowledge Environment, or MIKE. The goal of MIKE is to help LSE teachers, researchers, students, administrators, alumni, and visitors to cope better with the spiralling problem of information overload in their daily work. To do this we will:

- present them with a personalised desktop accessible both on and off campus, meeting their priorities and adapting to their changing needs
- link together the wide range of relevant information and administrative processes which are currently dispersed around the institution into a coherent and easily-searchable whole

- provide seamless access across a growing variety of licensed third party information content
- offer well-managed and well-packaged information to enhance the reputation, recruitment, and marketing of the LSE
- encourage content creators within the institution to make their material available online and integrate it into the information environment for easy access for authorised users

We believe that at LSE we are already well placed to create a successful MIKE because we have considerable experience and expertise across all our information services: Library, IT Services, Management Information Services (MIS), Website Services, and Centre for Learning Technology. And there are a number of building blocks already in place across these services which we can bring together now and enhance as a good seedbed for the MIKE. These include:

- LSEForYou, created by MIS and winner of a number of national and international prizes. This is a portal approach to providing administrative services so that students and staff can look up classroom bookings, fee payment status, salary payment status, and many more pieces of information from the corporate database [<http://www.lse.ac.uk/lseforyou/>]
- Electronic Library, one of the first services on the web to bring together a subject search approach to a large number of electronic sources, including access to over 5000 licensed full text e-journals and a wide range of free websites validated by subject librarians [<http://library-2.lse.ac.uk/EL/info/>]
- Online learning as part of a growing number of courses. The Centre for Learning Technology, using WebCT, is helping academic staff to embed online elements to most taught courses in the LSE, and through this and their work with the Library on electronic course packs they have a wealth of experience in the way learners use and interact with e-learning materials [<http://teaching.lse.ac.uk/tech/>]

A number of externally-funded projects can also be regarded as potential building blocks for MIKE, including:

- ANGEL, a Library project which has developed an understanding of the different roles that users have and methods of mediating access to information resources based on those roles. ANGEL has developed a Resource Manager software product which can be used to manage collections of electronic resources, and indeed the Electronic Library makes use of this product [<http://www.angel.ac.uk/>]
- DELIVER, a joint project between the Library and the Centre for Learning Technology which aims to develop integration tools to enable Virtual Learning Environment-based resources (in the case of the LSE this

is WebCT) and Library-based resources to be accessed easily from each other [<http://www.angel.ac.uk/DELIVER/>]

- SECURe, a collaborative project led by the Library but also involving staff from IT Services and Management Information Services, which is working towards an integrated system for controlling access to resources without multiple username and password challenges [<http://www.angel.ac.uk/secure/>]

Work for MIKE has started in two broad areas at the moment but still has a long way to go:

1. Content. A Content Group is identifying the large range of information resources required by members of the LSE community by holding focus group meetings with students, administrators, academics, and alumni. The next tasks will be to classify and prioritise the identified needs and design a metadata/content management structure for searching them. We will investigate existing products and schemes such as the proposed RSLP Collection Description Schema [1] and such proprietary products as Metalib [<http://www.aleph.co.il/MetaLib/>] and iPort [<http://oclc-pica.org/?id=106&ln=uk>]. The information content will remain in separate but linked repositories.
2. Technical architecture. The IT staff involved in MIKE will be responsible for selecting the portal software and for defining the ways in which different components of the system interact. They will also create the access management layer to connect the user at the desktop to the relevant content in one or more of the collection level registries, building on the SECURe Project. This will be achieved through single sign-on authentication and a sophisticated authorisation system which will deliver relevant resources only to users entitled to access them.

3 The portal

Most portal implementations in universities (including the current LSEForYou system) can be described by a generalised two-layer model, in which the portal or presentation layer provides the end-user interface, and is configured to access a number of content sources or collections of resources (Figure 1). This works well and is cost-effective to maintain when the number of content layer sources is relatively small, their structure and interfaces do not change frequently, and they are under local control. Most of the content to which access is enabled by LSEForYou resides in locally-managed Oracle® databases. Specialised user interfaces, rather than the general-purpose portal, are used for more complex transactions such as those required by administrative staff. For example, the SITS system is used at LSE for management of student record data, but students can access a restricted view of the data via LSEForYou.

Problems arise when this model is extended to include larger numbers of content collections, under more diverse and

autonomous management, and these problems may not be evident until attempts are made (as at LSE) to extend the

[<http://www.blackboard.com/>] may be used as a 'learning portal' to access course materials; and a library management system will be present for access to full-text electronic library materials and metadata (i.e. catalogue and stock management records) describing traditional printed library material. Indeed, most of the information systems in use can be described in a generalised way as a content store, specialised to suit the structure of the data involved; and a presentation layer or user interface, specialised to suit the kinds of transactions between the system and different types of users.

Usually it is necessary to duplicate some information across each of these discrete systems. For example, administrative systems will hold the primary or definitive records of all students and staff. These must be exported, imported and kept updated in a virtual learning environment to control access to course material. Similarly, student and staff records must be duplicated to a library system (often with the addition of records for other library users, who are not otherwise members of the university) as patron records, to manage lending and other functions. The university library often introduces additional problems in this area too, because there may be large numbers of authorised 'external' users of the library who are not otherwise known as 'members' of the university.

The principal reasons for a portal to need the authenticated identity of a user are to enable **personalisation** of portal content and functions, and to ensure access **authorisation** - i.e. that users are permitted access only to information for which they have rights, for reasons of confidentiality, security and economy.

Thus the two key areas on which a MIKE must focus are metadata about **resources**, and metadata about **users**. As well as enabling access with fewer discontinuities across diverse information resources, a MIKE must also impose **access management**.

4 The middleware: Collection Level Registries

The architecture for MIKE introduces a third, middleware layer between purpose-specific content management systems and presentation-layer user interfaces (Figure 2). The middleware components are a number of Collection Level Registries (CLRs) that provide a level of indirection between the (or, any number of) user interfaces or portals, and the content stores. This is similar in some ways to ideas being developed in the MIT Open Knowledge Initiative [<http://web.mit.edu/oki/>].

Separating the interfaces to different content sources from the portal itself brings a number of benefits:

1. Each administrative domain can continue to manage autonomously the content for which it is responsible, maintaining metadata about each content collection in its own CLR.

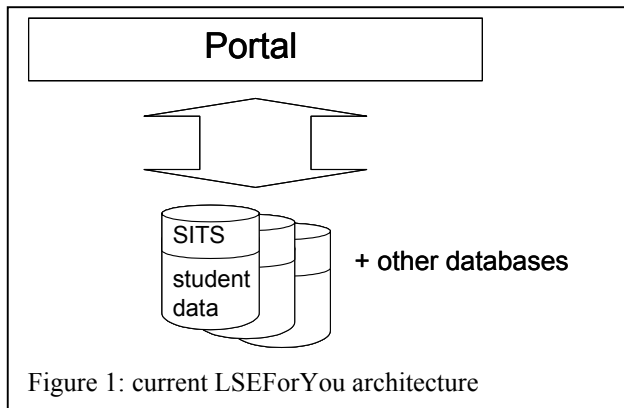


Figure 1: current LSEForYou architecture

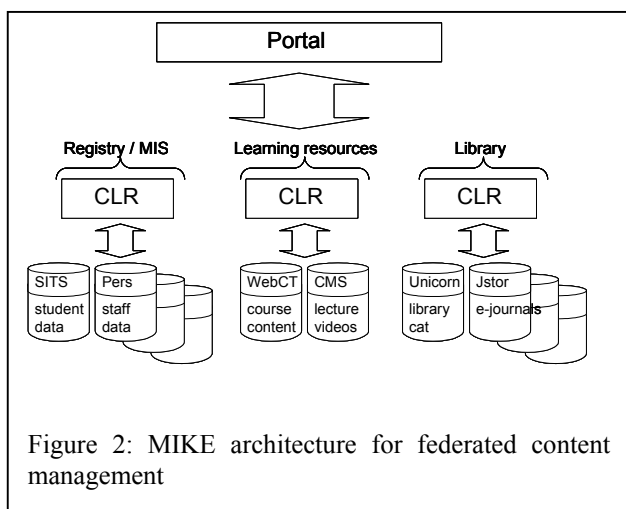
scope of the portal to encompass content sources managed by two or more distinct administrative domains of the university. Such problems for portal implementers are often 'caused' by the university library, which complicates a previously simple task by revealing the existence of hundreds (possibly thousands) of 'new' electronic resources, all dearly beloved by some important users, and all configured, managed and licensed-from a vast array of external suppliers who may, or may not, be willing to negotiate about the way in which their resources are accessed. The university is unlikely to have much control or influence over changes to these resources and the interfaces they expose, which will 'break' links from the portal. If system administration functions for the portal are the responsibility of, say, the IT support service, and a staff member in the library is responsible for liaison with a resource supplier, there will be additional barriers to awareness of an impending change of this sort, and the contacts and information needed to respond to the change.

Academic library consortia already have considerable experience in attempting to integrate collections of such autonomous content sources. For example, the InforM25 virtual union catalogue system [<http://www.m25lib.ac.uk/>] provides cross-searching of a large number of library catalogues that all nominally conform to the well-defined Z39.50 standard [<http://www.loc.gov/z3950/agency/>]. But diversity between implementations of the standard by the owners of different content stores (or the suppliers of their catalogue systems) renders the addition of each new source a far from trivial task. Furthermore, changes (such as routine software and configuration upgrades) made by those owners, which they may not even suspect will be evident at the portal, are often the cause of unexpected remedial work which must be done to re-make a broken connection.

The existing information environment of a university could be depicted (with a great deal of simplification) in a way that models typical organisational domains of the institution. Thus, one or more administrative user interfaces are provided to administrative data (such as student, staff and course/curriculum records); a virtual learning environment such as WebCT [<http://www.webct.com/>] or Blackboard

- More than one portal can be economically implemented and maintained. This may be trivially useful where ‘live’ and test versions of the portal must co-exist; but it may also be productive to experiment with several radically different portals, intended for use by different groups of users such as students, researchers or alumni.
- The architecture is scalable to include very large numbers of new content collections; if necessary, by adding CLRs to balance the loading.

Each CLR must, of course conform to agreed standards for machine-to-machine interfaces with the portal(s) or presentation layer for which it provides services.



Although a content collection can be described in the CLR using an accepted schema such as the RLSP Collection Level Description, further metadata is needed for the architecture to operate. Our understanding of what is required has been developed through work on a number of national projects such as HeadLine [<http://www.headline.ac.uk/>], which implemented such a schema as a relational database, accessed by presentation-layer applications using SQL. The HeadLine Resource Datamodel [2,3] included all the attributes that would be recognised as a bibliographic description, but added metadata to describe how, technically, a resource collection could be searched and content items retrieved and how end-user access to the collection was controlled, anticipating in some ways the devolved authentication schemes that are now being implemented by projects such as AthensDA [http://www.athensams.net/development/devolved_authentication/] and Shibboleth [<http://shibboleth.internet2.edu/>].

5 The middleware: Content management

The ANGEL Project [<http://www.angel.ac.uk/>] developed this concept further, implementing a Resource Manager (which is available as Open Source software) that encapsulated the relational database within an XML-based messaging protocol for requests and responses. This enables

details of the resource collections to be accessed by portals (or any presentation-layer application) in other commonly-used formats, such as the Open Archives Initiative Protocol for Metadata Harvesting [4] and Z39.50 [5]. The ANGEL Resource Manager is designed in such a way that new access protocols can be added relatively easily. For example, RDF Site Summary (RSS) [<http://web.resource.org/rss/1.0/>] is commonly used by many portal products to enable ‘news feeds’; but RSS is not inherently tied to this purpose, and is well-suited for rendering brief, linked descriptions of resources, rather like the results of a Web search-engine. So a RSS feed was implemented in the Resource Manager to make library resources easily accessible via portals.

We do not take any particular ideological stance on the selection of appropriate software components to implement elements of the MIKE; neither towards a completely in-house (or Open Source) development; nor in favour of total ‘outsourcing’ using proprietary commercially supported products. We will inevitably need to make use of many existing ‘legacy’ systems in the early stages, and the architecture will need to accommodate and include many systems that are required in the future. Instead, the approach we take is to define standards for inter-operation and the interfaces between major components. In the short-term this will incur penalties of lower runtime efficiency (than the alternative of a more tightly-integrated single proprietary solution); but we believe these will be soon outweighed by the greater flexibility and scalability of a high-level architecture over which we have greater control; and that runtime performance, where it poses problems, can be easily improved with more hardware and network capacity. This is relatively very much cheaper than dealing with the maintenance and periodic ‘big bang’ replacement of a monolithic solution.

As an example of how this ‘mixed economy’ approach might be implemented, commercial products such as MetaLib/SFX or iPort could be used to implement one (or more) of the CLR components. MetaLib/SFX could co-exist, in the MIKE architecture, with an Open Source product such as the ANGEL Resource Manager, both using the published OpenURL standard [6] to provide client applications with a ‘deep link’ to items such as individual e-journal articles, in managed content collections that may be in-house repositories or (more likely) the remote holdings of a publisher, accessed under license.

6 Access management: the SECURE Project

In this field as in others, LSE has been extremely active in making and developing relationships with other Higher Education institutions, standards bodies and research programmes; in the UK, Europe, USA, Australasia and other parts of the world. It is clear that no university can or should attempt to be a self-reliant or self-contained information environment, and the design goals of MIKE are that it should operate (in most respects) regardless of the physical boundaries or network topology of our campus, and of the

'regulatory boundary' defining resources over which we have administrative control. MIKE should also address the problems of authorised LSE users who want access from anywhere (in the world), at any time.

However, there are many boundaries that it must enforce - ensuring that only authorised users have access, and enforcing the restrictions properly required by considerations of intellectual property, privacy, security and plain economics [7].

The access management functions of MIKE are supported by the SECURE (Secure Environment for Certificated Use of Resources) Project. SECURE will implement and test a full-scale solution, portable and scaleable to other Further & Higher Education institutional environments, to the key problem of achieving fine-grain authorisation of access by all users in an institution to the complete spectrum of relevant information resources. The range of resources must include those under a high degree of institutional control, with relatively high security requirements (such as sensitive staff and student personal data), through to such out-of-domain resources (over which the institution has no direct control) as subscription e-journals to which the institutional library negotiates licensed access for users.

In particular, SECURE will use the existing and emerging technologies of Public Key Infrastructure (PKI) with X.509 digital certificates [<http://www.ietf.org/html.charters/pkix-charter.html>] and Shibboleth [<http://shibboleth.internet2.edu/>] to implement the necessary authentication infrastructure to support this, on as large a scale as project and institutional resources allow. To meet the practical needs of MIKE, this will involve all staff and student members of LSE (approximately 10,000 individuals). This phase of SECURE will focus on documenting and solving the management problems raised by institutional scale deployment of PKI and Certificate Authority services, including appropriate strategies for certificate issue, revocation, and linking the authorisation of certificates with existing registration processes that establish real-world identities of students, staff and other users of campus IT services and managed information resources.

7 The need for interim solutions

The vision for LSE described earlier in this paper will be realised at some time in the future and whilst some elements are already implemented, others are clearly some way off yet. But the problems and issues that the MIKE seeks to address are real problems that are presenting themselves within many universities today. The answers cannot wait until some unknown date when all will become possible; practical, pragmatic solutions that can be deployed now are needed to fill the gap.

It would be foolish, however, to disregard the vision and to implement solutions today that do not fit our model of the future. These stop-gap systems and services must take all they can from the MIKE architecture and must only be creative

where necessary. It is also preferable that, as the vision is realised, these systems and services will evolve into the MIKE architecture rather than being short term systems in their own right. This has a significant impact on the design and development process. It is more acceptable to deploy a system where some aspects are less than ideal if it is known that those elements will be replaced in six or twelve months time.

There are, of course, many others but the British Library of Political and Economic Science at LSE provides a good example of the problems that universities face today. This major research library serves a much wider community than just the staff and students at LSE. It houses many very important collections which are used by scholars from across the world. It is also a valuable research resource which is used by students, commercial users and members of the general public. In the past, it was necessary to make the journey but once there, all the resources were available in physical printed form. It may have been an enormous task, but collecting these resources, classifying them and making them available to those who wished to read them was at least possible.

The age of electronic information has brought about significant changes though. The cost of purchasing and handling printed media has led a massive drive towards the provision of electronic only versions, particularly of scholarly journals. Increasingly, important academic work is never published in the traditional manner. This clearly has an impact on the professor from China who has made the trip to LSE, as he can physically access the printed material but he now needs a computer to be able to see the whole collection.

A brief application of the MIKE vision and architecture to this problem will demonstrate that in the future this will be a trivial matter. The professor from China will identify himself as such and will be challenged for his authentication token which will then be verified with his institution or some other competent body. The portal will welcome him as a visitor to the LSE and will tailor his view to the resources he is most likely to use. Access to electronic resources will be then be mediated by the owners of those resources and allowed either of the basis of his own identity or through being a visitor to the LSE. If all the information required is held in electronic form, it will be possible to be authenticated as a virtual visitor to the LSE and our professor need not even make the trip.

However, today there are many challenges which must be understood and overcome before any large scale, international scheme which provides this type of facility will be implemented. LSE, like most institutions, requires users to identify themselves and to be authenticated before they can use a networked computer, but how can this process be managed for users who are not members of our institution? The suppliers of electronic information are also anxious to protect their revenue streams and will need to change the way that they control authentication and authorisation if they are to meet the demands for ubiquitous access to their products. These resources are provided under a myriad of different

licenses, each with its own terms and conditions as to what classes of user are permitted, which makes it very difficult to determine who is entitled access to what. The success of the Athens service in the UK has shown that some are willing to entrust a third party with the information to verify authentication; many however are not.

8 Visitor access: the UK Computing Plus Project

The difficulty for visitors to our libraries threatens to become an acute problem for the UK Libraries Plus scheme which has been of great benefit to part time students and distance learners within the UK higher education community for some years. Under the scheme, eligible students are able to register to use the library facilities at up to three other member institutions with students generally choosing those nearest where they live or work. Once registered, users have access to printed reference material and also limited borrowing rights but do not have access to resources that were available in an electronic format only. The extension of the scheme to cover access to IT facilities and to electronic resources would further enhance its value. In order to explore these challenges, six pilot projects were commissioned, each with a different approach to tackling these issues. One of these pilot projects, which are collectively known as UK Computing Plus, is at LSE.

In considering what model should be adopted for UK Computing Plus and in the absence of a common, national scheme for authentication and authorisation, the LSE philosophy has been to make each institution responsible for the administration and access management of its own students. It was also felt that participants in the scheme should be able to access the same resources as a "standard" student and should be neither advantaged nor disadvantaged by physically being in another institution. These two factors have influenced the model developed.

The joint goals are achieved through the creation of a virtual extension to the 'home' institution. As far as possible, the visiting student is in the same position as if they were sat at a workstation in their own institution and so has access to those resources which are managed by (or, to which access is licensed by) their 'home' institution.

There are a number of different technologies available which could be used to implement this service, all of which require the home institution to provide an appropriate server and the receiving institution to provide client software. Many institutions already had such an infrastructure in place, based on either web proxy servers or thin client technology (such as Citrix) in order to support staff and students who work from home. The receiving institution will need client configuration information for each user and some work is required to develop a simple but secure user interface for the pilot scheme.

The operation of the system is very simple. When a UK Libraries Plus student registers to use the library at LSE, they are given the UK Computing Plus username and password. The same username and password are used by all users under the scheme. When a user logs onto an LSE computer with this username and password a program is started which requires them to select their home institution from a menu. The security policy is set so that should they ever exit this program, the session is ended and they are logged out. Having selected their institution, one or both of the following will happen:

1. A web browser window is opened which is pre-configured to route all traffic through their home institution web proxy server. This proxy server should require them to authenticate with their local (i.e. home institution) username and password.
2. A Windows Terminal Server session is opened connected to their home institution's terminal server. Again this should require them to authenticate with their local username and password.

Note that, other than the generic username and password to access the menu program, all username and password challenges are for their home institution credentials. Once a national scheme for authentication is in place, the need to connect to a server at the home institution will disappear as the user's identity credentials will be verified through the national scheme, (although in many instances users will still want to be connected to their home server).

It should also be noted that the user does not have to know anything about how they connect to their home institution. A database has been gathered of the configuration details required for these connections. For this project this is held at LSE and collecting this information has, in fact, proved one of the most difficult parts of the whole project. In the MIKE architecture, this database becomes an item in a Collection Level Registry, (CLR), initially held at LSE but it could move to a national CLR should that prove more appropriate.

One drawback of the project, as implemented at the moment, is that the control of access to resources is still based solely on the basis of the user's home institution. Whilst this is still of benefit to the distance learners who may live too far away to have physical access to a computer on their "home" campus, as well as those who wish to work with a combination of printed and electronic sources, for many the value is limited. The real value will come only when global standards have been implemented for the mediation of access to electronic resources.

The UK Computing Plus initiative is specifically for part-time students and distance learners and so the LSE pilot service is focused on this particular group. However, as stated earlier, the LSE Library's role as a national and international research library brings in many other types of visitor and the model developed for the pilot is suitable for use by any group of visitors where there is an equivalent concept to the 'home' institution. Access to the system has already been extended to

other visitors with borrowing rights from UK institutions and further extensions are planned. If a national PKI for the UK population were ever implemented it might even be possible to extend the scheme to the general public.

Acknowledgements

The ANGEL, SECURE and DELIVER projects have all been part-funded by JISC (the Joint Information Systems Committee of UK Further & Higher Education funding bodies). The UK Computing Plus scheme is supported by SCONUL (the Society of College, National & University Libraries) and UCISA (Universities & Colleges Informations Systems Association). All proprietary trademarks used are acknowledged.

References

- [1] Powell, Andy et al. "RSLP Collection Description", *D-Lib*, volume 6 number 9, DOI: 10.1045/september2000-powell
 - [2] Graham, Stephen. "The HeadLine Resource Data Model", *VINE*, issue 117, pp 13-17.
 - [3] Paschoud, John. "The filling in the PIE - HeadLine's Resource Data Model", *Ariadne*, issue 27, <http://www.ariadne.ac.uk/issue27/paschoud/>
 - [4] Open Archives Initiative. "Protocol for Metadata Harvesting", <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
 - [5] Library of Congress. "Z39.50 Maintenance Agency Page", <http://www.loc.gov/z3950/agency/>
 - [6] National Information Standards Organization. "The OpenURL Framework for Context-Sensitive Services", http://www.niso.org/committees/committee_ax.html
 - [7] Paschoud, John. "All Users are Not Created Equal! - How to decide Who Gets What from your Hybrid Library", *Internet Librarian International* (1999), <http://www.headline.ac.uk/public/diss/index.html#90330JP>
-