Digital Tools for Library and Information Studies

A Methods Network Working Paper

Defining discreet scholarly territories for all disciplines is problematic but it could be argued that Library and Information Studies (LIS) is more problematic than most when it comes to understanding the scope of its remit. It addresses issues relevant to every other academic subject and demands the engagement, to varying degrees, of almost everyone whose task it is to carry out research. Wherever information is aggregated on a scale that renders it difficult to navigate, retrieve, analyse, preserve or store that data, opportunities arise for librarians and information specialists to intervene. In addition to pan-disciplinary involvement, LIS also addresses the management of all types of data encompassing text, sound, still and moving images, resulting in a community of practitioners that work in an enormous variety of environments.

Bearing in mind the above, it seems sensible to state some objectives for this paper and to arrive at some definition of its scope and ambition. The principal aim is to provide an overview of *some* of the options that are available to academics who wish to embrace the use of digital tools when engaging in the sort of arts and humanities related research that principally involves issues to do with information management. The selection of tools featured will reference the discussions that have taken place as a result of Methods Network activities and may also reiterate the content of other working papers in this series, all of which share this document's broad interpretation of the term 'tools'. The intention is that the Methods Network will use this paper as a catalyst for discussion between all individuals who have an interest in LIS related issues and the expectation is that their input will add significant value to a document that is, by definition, in draft format.

Until a more formal forum for discussion is developed, please feel free to send any comments, additions, suggestions and corrections to neil.grindley@kcl.ac.uk.

LIS Research

For the purposes of this paper, Library and Information Studies should be understood as representative of a group of activities that are referred to by a number of different appellations. Whilst acknowledging that 'information science', 'information systems', 'information management', 'information and library studies', 'librarianship' (and many other combinations of terminology) all bring something different to the field,[1] it is convenient in this context to homogenize this diversity. Whilst simplistic, it would seem to be the only concise way of beginning to tackle such a complex area of research. This complexity is illustrated as soon as one contemplates the nature of the relationship between LIS research and the notion of 'information'. On the one hand, there are many instances where research is very much focused on the divisible, categorical and meaningful interpretation of data.[2] A common example of this is the construction or refinement of classification schemes and ontology models which is a routine activity within LIS. On the other hand, other areas of research are unconcerned with 'meaning' and are more focused on the formal properties of 'information'. Taking this approach, questions might be asked such as: how does information flow?; where does it flow?; why doesn't it go where it's needed when it's needed?; why is there so much of it?; do we need it all?; if we don't, how do we get rid of it?; how do we choose what to get rid of?; and so on and so forth.

When contemplating the amount of discreet theory that is attributable to the discipline of information studies, Tom Stonier proposes a sample question, 'How much information is contained within a steam engine?',[3] which declares an intention to deal with information in a rigorous, scientific and quantitative way.

[1] For a list of schools, departments and courses in the UK relating to Information Studies, see: http://informationr.net/wl/wlist2.html
[2] For a discussion of 'data' in relation to 'information', 'knowledge' and 'wisdom', see: Hey (2004)
[3] Cited by Webber (2003)

As a position, it self-consciously puts distance between itself and branches of 'soft' LIS research[4] that focus on qualitative analysis, e.g. reactions to interface design. This acknowledgement of the breadth and complexity of the discipline is an attempt to manage expectations in respect to the eclectic nature of the tools referred to during the rest of this paper, and the very broad way in which some of them are dealt with.

Starting with the most obvious and publicly visible activity that relates to library work, it may be useful to acknowledge the pivotal role of digital tools in the realm where theory meets practice and where the user interacts with the information specialist, i.e. at the issue desk. The use of barcode scanning equipment and library management software such as Aleph[5], are a standard feature in almost all libraries and as such, are of enormous significance within the discipline. Whilst acknowledging initiatives such as the recent use of RFID (radio frequency identification) for the tagging of library objects,[6] and the ongoing changes to the architecture of large and sophisticated library management systems, it might be useful to look beyond these well understood processes to focus momentarily on tools that address the demands of fetching and retrieving physical resources, before moving onto more influential and representative areas of LIS research.

The *UJI Librarian Robot*, developed by the Robotic Intelligence Lab at the Universitat Jaume I (UJI, Spain)[7] is an attempt to automate the process of shelving and reshelving and uses sensors and robotic arms to identify and select bar-coded objects situated on library shelves. (see fig.1)

Fig. 1 The UJI Librarian Robot[8]

Whilst this system is a novel and rather marginal example of tools usage, the appropriation of knowledge and expertise from both robotics, engineering and artificial intelligence is very significant in terms of cross-disciplinary collaboration and is an area of research that could be applied using other programmable devices, for instance: conveyor belt technologies; automated moving stacks; mechanized provision of short loan items; etc.

The implementation of intelligent automation is clearly a sensible and desirable division of labour where machines can undertake repetitive manual tasks as well or more effectively than humans.[9] Where this is possible to implement, information specialists are then free to devote time to the sort of intellectual processes that more accurately define library-related work, principal amongst which are:

- the creation and application of organisational systems;
- the implementation of standards-based information management principles

[4] For a discussion of 'hard/soft/pure/applied' classification of academic disciplines, see Biglan (1973)
[5] Ex Libris, Aleph, http://www.exlibrisgroup.com/aleph.htm, (accessed 22 May 2007)
[6] For an account, see:CILIP, Implementing RFID at Leeds Met – Helen Loughran and Dilys Young, http://www.cilip.org.uk/publications/updatemagazine/archive/archive2007/april/loughranandyoung.htm, (accessed 22 May 2007)
[7] Robotic Intelligence Lab, UJI, Spain, http://www.robot.uji.es/, (accessed 11 May 2007)
[8] Images taken from: Robot Gossip, Robot Librarian, http://robotgossip.blogspot.com/2005/10/robot-librarian.html, (accessed 11 May 2007)
[9] For a discussion of this, see, Norman (1999)

- the development of sophisticated information retrieval techniques
- the application of measures to ensure usability and accessibility of materials

Information Systems and Digital Libraries

As Besser[10] points out, if a large collection of books and associated material has all been assembled in one location, it doesn't necessarily mean that it should be called a 'library'. To concur with this definition, he argues that skilled intervention is required to arrange and manage the material in such a way that it can be stored and retrieved in the most effective manner possible, and this logic can also be used as a basis for defining how objects should be handled within digital libraries. Before addressing this issue however, it is useful to remember that the concept of an object being anything other than something that is *referenced* by a library system (because of its analogue nature) is relatively recent, and it is the incremental developments to older systems, some with analogue origins, that have provided information scientists with the means of effectively delivering the current generation of enormous and comprehensive catalogues.

Classification systems such as *Dewey*,[11] *Universal Decimal* and *Library of Congress* have been in existence since the end of the nineteenth century and their hierarchical structures and alphanumeric strings were ideally suited to inclusion within computer systems. National variations of the *MARC*[12] (Machine Readable Cataloguing) format, first developed in the 1960's, are now used by libraries all over the world for defining bibliographic and related material in their catalogues, thereby greatly easing the transferral and exchange of records between organisations. A relatively recent iteration of the format is called *MARC21* and is a combination of the U.S and Canadian versions, based on the ANSI (American National Standards Institute) standard *Z39.2*.[13] As of June 2004, the British Library adopted *MARC21* as its cataloguing format in the course of implementing a new integrated library system.[14]

The use of classification systems and databases containing vast numbers of bibliographic records, in some cases joined together into very large union catalogues, (e.g. the *Copac Academic and National Library Catalogue*,[15] and *SUNCat* – the Serials Union Catalogue[16]) is the accepted and expected method by which users access library indexes. These *OPAC* (Online Public Access Catalogue) systems are inevitably now delivered as web-based interfaces and as such are capable of visualizing results in much more flexible and user-configurable ways than before.

Focusing specifically on the creation and development of digital libraries (i.e. systems that store items as digital objects and facilitate delivery of those objects to users via computerised means) there are currently various platforms which organisations are using to manage and deliver resources at departmental or institutional levels.
The development of Digital Library systems such as *DSpace*,[17] *EPrints*,[18] *Fedora*,[19] *Greenstone*[20] and *dLibra*[21] is an attempt to combat the sort of *ad hoc* practices, rife in all organisations, where users who are

[10] Besser (2004)
[11] Online Computer Library Center (OCLC), Dewey, http://www.oclc.org/dewey/, (accessed 11 May 2007)
[12] Library of Congress, MARC Standards, http://www.loc.gov/marc/, (accessed 12 May 2007)
[13] ANSI, Z39.2, http://webstore.ansi.org/ansidocstore/product.asp?sku=ANSI%2FNISO+Z39.2-1994, (accessed 11 May 2007)
[14] British Library, http://www.bl.uk/services/bibliographic/nbsils.html, (accessed 12 May 2007)
[15] Copac, http://copac.ac.uk/about/, (accessed 12 May 2007)
[16] EDINA, SUNCat, http://www.suncat.ac.uk/, (accessed 13 May 2007)
[17] DSpace, http://www.dspace.org/, (accessed 18 May 2007)
[18] EPrints, http://www.eprints.org/, (accessed 18 May 2007)
[19] Fedora, http://www.fedora.info/, (accessed 18 May 2007)
[20] Greenstone Digital Library Software, http://www.greenstone.org/cgi-bin/library, (accessed 18 May 2007)
[21] Poznan Supercomputing and Networking Center, DLibra, http://www.greenstone.org/cgi-bin/library, (accessed 18 May 2007)

uncertain what to do with the files they acquire or create often assign them to oblivion on institutional file or mail servers in deep and unmanaged folder structures.

Taking one of these as an example, *DSpace* is a digital repository system jointly developed by MIT Libraries and Hewlett-Packard Labs that 'captures, stores, indexes, preserves, and redistributes an organization's research data.'[22] It is a freely available open source software platform that accommodates all forms of digital material including text, image, audio and video files and handles the submission and re-delivery of those materials via a web interface. In addition to the user-friendly interface, *DSpace* offers the following benefits for those choosing to deposit materials using this tool.

- The system formalises the requirement for metadata to be added to objects, allowing them to be indexed for browsing and searching and defined within groups of similar objects that are placed together in logical collections.
- Groups are designated as belonging to a particular *community* which can correspond to specific parts of an organisation (e.g. department, research centre or school)
- Communities are modular and can be extended across institutional boundaries
- Locations for objects can be specified with persistent URL names allowing for sustainable and reliable referencing
- Preservation of file formats is managed by automatically updating material to avoid format and technology obsolescence[23]

The functionality referred to above is indicative of the features that other digital repository software systems may offer and implementations of three of these systems have either been carried out or are planned in a number of UK Higher Education organisations, for instance:

DSpace – Loughborough University - https://dspace.lboro.ac.uk/dspace/
e-Prints – University of Southampton - http://eprints.soton.ac.uk/
Fedora – University of Hull - http://www.fedora.info/wiki/index.php/University_of_Hull

The planning required to set up these systems is significant and may often be contemplated in conjunction with complementary systems. At the University of Hull, the browser front end that serves as a gateway into their system is facilitated by *uPortal,*[24] a free open-standard collaborative portal tool which allows users to customize their access to the institution's online resources. This organisation is also looking at implementing the *Sakai*[25] Collaboration and Learning Environment, another free and open source product, which is maintained by a largely U.S. based community of developers and users. A set of generic collaboration tools forms the core set of features within the system (e.g. wiki, forum, chat room, WebDAV , RSS, scheduling), but it is extensible to allow additional tools to be incorporated. Lancaster University Centre for e-Science is currently using a *Sakai* environment for its *ReDress* project,[26] whose remit is to raise awareness of e-Science related technology initiatives amongst social scientists. One of the features of this site is a highly effective online archive of seminar and workshop presentations featuring full audio and video versions of the papers synchronized with their related *Powerpoint* slides.[27]

The EU funded DELOS project[28] (Network of Excellence on Digital Libraries) is an initiative to integrate and coordinate the ongoing work of major European efforts in the field of Digital Libraries and is currently

[22] DSpace, http://dspace.org/introduction/index.html, (accessed 12 May 2007)
[23] A visual description of the DSpace system is available at: DSpace, http://dspace.org/introduction/dspace-diagram.pdf, (accessed 12 May 2007)
[24] uPortal, http://www.uportal.org/index.html, (accessed 13 May 2007)
[25] Sakai, http://sakaiproject.org/, (accessed 13 May 2007)
[26] Lancaster University, Centre for e-Science, ReDress, http://redress.lancs.ac.uk/, (accessed 13 May 2007)
[27] Lancaster University, Centre for e-Science, ReDress Presentations, http://redress.lancs.ac.uk/Workshops/Presentations.html, (accessed 13 May 2007)
[28] DELOS, http://www.delos.info/, (accessed 13 May 2007)

working on the development of a *Digital Library Reference Model*[29] and a prototype implementation of a *Digital Library Management System*.[30] This ambitious collaborative project seeks to define how the next generation of digital libraries will be built and used and incorporates a middleware environment called *OSIRIS* (Open Service Infrastructure for Reliable and Integrated Process Support) and a set of services referred to as *ISIS* (Interactive Similarity Search). These components and the applications and central functions that they translate between are expected to be able to support a huge range of functions covering the entire data life-cycle from acquisition, through retrieval and analysis (for all data types including automatic feature extraction from image and audio material), through to archival and long-term preservation.

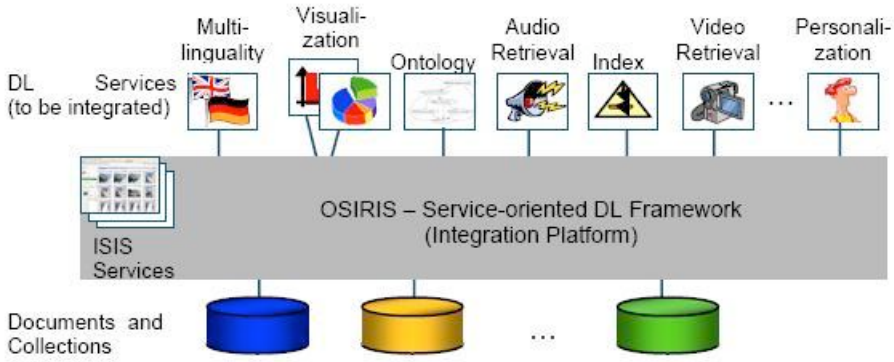

Fig. 2 Representation of the *DelosDLMS*[31]

Such a proliferation of function is symptomatic of the fact that the term 'digital library' defines an entity which requires attention to strategy on many fronts, not least on issues to do with data standards and information formats. One of a number of useful papers by the Arts and Humanities Data Service (AHDS)[32] introduces issues relating to digital repository systems and refers to their adoption of the *OAIS* (Open Archival Information System) reference model.[33] Originating from work done in scientific disciplines, this is a framework that expresses concepts relating to the long term preservation of, and access to, digital information. This paper also makes useful reference to the *Storage Resource Broker* (SRB) data management tool which enables data that is distributed across multiple storage devices to be viewed as if it were part of a single file system.[34]

Data Management

Though closely entwined with the previous section, focusing on 'data management' issues affords opportunities to concentrate more specifically on data standards, ontologies, thesauri, data types and data processing methods. The Open Archives Initiative (OAI – acronymically similar but conceptually distinct from OAIS[35]) is an influential community of individuals and organisations that are interested in promoting efficient methods for broadly disseminating information using interoperable standards. It is formally

[29] DELOS, A reference model for Digital Library Management Systems, http://www.delos.info/index.php?option=com_content&task=view&id=345&Itemid=, (accessed 13 May 2007)
[30] DELOS, DelosDLMS: Global Prototype Development, http://www.delos.info/index.php?option=com_content&task=view&id=502&Itemid=, (accessed 13 May 2007)
[31] DELOS, DelosDLMS: Global Prototype Development, http://www.delos.info/index.php?option=com_content&task=view&id=502&Itemid=, (accessed 13 May 2007)
[32] AHDS, Curl Case Study, http://ahds.ac.uk/preservation/curl/curl_case_study_d4.pdf, (accessed 13 May 2007)
[33] Research Libraries Group (RLG), OAIS Activities, http://www.rlg.org/en/page.php?Page_ID=3201, (accessed 13 May 2007)
[34] San Diego Supercomputing Center, SRB, http://www.sdsc.edu/srb/index.php/Main_Page, (accessed 13 May 2007)
[35] D-Lib, OAI and OAIS: What's in a name?, http://www.dlib.org/dlib/april01/04editorial.html, (accessed 14 May 2007)

supported by the Andrew W. Mellon Foundation,[36] the Coalition for Networked Information,[37] The Digital Library Federation[38] and the National Science Foundation,[39] but also has non-U.S. based organisations taking an active part, notably the Joint Information Systems Committee (JISC) and UKOLN (UK Office for Library Networking) who were both represented at a meeting in April 2006 which addressed the theme, 'Augmenting Interoperability Across Scholarly Repositories'.[40]

One of the principal concepts promoted by OAI is the use of the *OAI-PMH* (Protocol for Metadata Harvesting),[41] a method by which *XML*-based descriptive metadata records can be automatically acquired by *OAI-PMH* compliant systems. Typically records would take advantage of well established methods such as using *Dublin Core*[42] or *MARCXML*[43] format data. The OAI site lists around thirty different tools[44] that have been developed by members of the OAI community that relate to the protocol, examples of which include,

- *Archimede* - Archimede is an open-source software for institutional repositories. It features full text searching, multiplatform support, Web user interface, and more. Archimede fully supports OAI-PMH requests version 2.0.

- *OAI-perl Library* - A library of PERL language classes that allow the rapid deployment of an OAI compatible interface to an existing web server/database

- *ZMARCO* - ZMARCO is an Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) 2.0 compliant data provider. The 'Z' in ZMARCO stands for Z39.50; 'MARC' stands for MAchine-Readable Cataloging; and the 'O' stands for OAI, as in the Open Archives Initiative. ZMARCO allows MARC records which are already available through a Z39.50 server to relatively easily be made available via the OAI-PMH.[45]

Making data available using the above methods is a well established process and is in accordance with the recommendations given for content providers wishing to make their systems interoperable with the JISC Information Environment (JISC-IE).[46]

Whilst methods of standardising metadata and the use of protocols might appear to stretch definitions of 'tools' usage, it is difficult to not to refer to these initiatives when reflecting on LIS research territories, central as they are to the concerns of data management. A similarly critical area of research focuses on the development and application of ontologies, a term that overlaps to some extent with the related activities of glossary, controlled vocabulary, thesaurus and taxonomy creation. The purpose of an ontology is to map all of the objects or concepts relating to a field of knowledge into a systematic arrangement that then displays the relationships that exist between them and their properties in relation to the whole domain. The application of ontologies allows systems to operate more knowledgeably by allowing related semantic concepts to be aggregated together for search and retrieval purposes.

---

[36] Andrew W. Mellon Foundation, http://www.mellon.org/, (accessed 14 May 2007)
[37] Coalition for Networked Information, http://www.cni.org/, (accessed 14 May 2007)
[38] Digital Library Federation, http://www.diglib.org/, (accessed 14 May 2007)
[39] National Science Foundation, http://www.nsf.gov/, (accessed 14 May 2007)
[40] The report, hosted on the Andrew Mellon Foundation website but linked to from the OAI site is at: Andrew W. Mellon Foundation, http://msc.mellon.org/Meetings/Interop/FinalReport, (accessed 14 May 2007)
[41] OAI, http://www.openarchives.org/OAI/openarchivesprotocol.html, (accessed 14 May 2007)
[42] Dublin Core Metadata Initiative, http://dublincore.org/, (accessed 14 May 2007)
[43] Library of Congress, MARC21 XML Schema, http://www.loc.gov/standards/marcxml/, (accessed 14 May 2007)
[44] OAI, PMH Tools, http://www.openarchives.org/pmh/tools/tools.php, (accessed 14 May 2007)
[45] All three examples are taken from OAI, PMH Tools, http://www.openarchives.org/pmh/tools/tools.php, (accessed 14 May 2007)
[46] Ariadne, 5 Step Guide to Becoming a Content Provider in the JISC Information Environment, http://www.ariadne.ac.uk/issue33/info-environment/, (accessed 14 May 2007)

Data modelling exercises of this type might vary in size depending on the scope of the domain in question. So called 'upper' ontologies map downwards from very high level general concepts and examples of this include: *SUMO*[47] (Suggested Upper merged Ontology) which is limited to meta, generic, abstract and philosophical concepts; and *Cyc*,[48] a manually entered knowledegebase of more than a million human assertions formalising common-sense notions such as 'The earth orbits the Sun'. At a more specific level, ontologies exist for much narrower domains (e.g. Atomic elements, biological viruses) but they also vary in their complexity and in their applicability to be defined as ontologies. The *Dublin Core* schema is an example of a simple ontology, whilst *WordNet*,[49] a large lexical database of English with words grouped into conceptually related sets of synonyms, might alternatively be described as a semantic lexicon or a combination of taxonomy[50] and controlled vocabulary.[51]

Another influential ontology model which is widely used in the Museums and Cultural heritage sector is the *CIDOC-CRM*.[52] Full reference will be made to this in a forthcoming Methods Network working paper (see Digital Tools for Museums and Cultural Heritage Research), but to summarise, the CRM provides a common and extensible semantic framework that was originally developed for cultural information but is potentially of conceptual use to any domain of activity. The *FRBR* model (Functional Requirements for Bibliographic Records), developed by IFLA (International Federation of Library Associations and institutions) in 1998, took a fresh look at what functions bibliographic records perform and then systematically tried to map the bibliographic realm by means of defining entities, relationships and attributes.

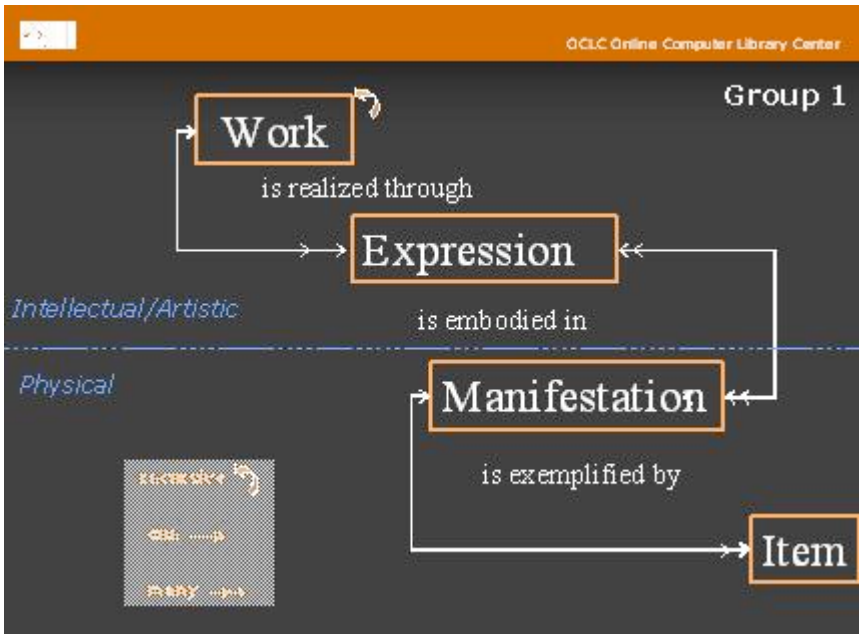

Fig. 3 The FRBR Group 1 bibliographic entities[53]

[47] SUMO, http://www.ontologyportal.org/, (accessed 14 May 2007)
[48] CyCorp, http://www.cyc.com/, (accessed 14 May 2007)
[49] Princeton University, WordNet, http://wordnet.princeton.edu/, (accessed 14 May 2007)
[50] Taxonomies only have parent-child relationships and cannot express more complex relations
[51] For a discussion of the uses of knowledge representation techniques in Linguistics, see: Methods Network, 'Tools for Linguistics' Working Paper, URL forthcoming ...
[52] The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC), Conceptual Reference Model (CRM), http://cidoc.ics.forth.gr/, (accessed 14 May 2007)
[53] OCLC, Overview and History of FRBR, http://www.oclc.org/research/presentations/childress/20050721-coasis&t.ppt, (accessed 14 May 2007)

The *FRBR* also considered what the users of catalogues 'do', and what constitutes the product of artistic endeavour.

More recently, a working group has been convened to look into the harmonisation of the *CIDOC-CRM* and *FRBR* and has concluded that there are mutually beneficial components in both models. The *FRBR* enriches the *CIDOC-CRM* with its notions of the stages of intellectual creation (see fig. 3), and conversely, the *FRBR* is able to appropriate a general model of historical events from the *CIDOC-CRM*.[54]

Discussion of data management issues inevitably requires reference to *XML* (and its related tools and techniques) and versions of the *CIDOC-CRM* model have been encoded in *RDFS* (Resource Description Framework Schema). This is based on *XML* and is an extension of *RDF*, a standard and universally machine-readable way of expressing information about web resources (e.g title, author, modification date, contents, etc).[55] *RDF Schema* provides a framework to describe application-specific classes and properties and has some resemblance to the use of classes and sub-classes in object-oriented programming, enabling hierarchical descriptions of objects to be encoded, eg. 'horse' can be defined as a sub-class of 'animal'.

```
<?xml version="1.0"?>

<rdf:RDF
xmlns:rdf= "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base=  "http://www.animals.fake/animals#">

<rdfs:Class rdf:ID="animal" />

<rdfs:Class rdf:ID="horse">
  <rdfs:subClassOf rdf:resource="#animal"/>
</rdfs:Class>

</rdf:RDF>
```

Fig. 4 Example of RDFS class (taken from W3 Schools tutorial)[56]

This clearly has applications to the construction of ontologies and the standard way of doing this with *XML* is by using *OWL* (Web Ontology Language), a W3C recommended tool that allows machine-readable semantic content to be processed and exchanged between different operating systems and application languages. *OWL* and *RDF* share the same look and feel but the former is defined with a larger vocabulary and has a stronger syntax, meaning that it has greater machine interpretability. It includes three sublanguages that sequentially increase in complexity and are nested in terms of functionality:

- OWL lite,
- OWL DL (includes OWL lite)
- OWL Full (includes OWL DL).

The use of various applications of *XML* for a wide variety of data management tasks is too broad a field to summarise adequately in this paper but the following approaches are indicative of ways that the use of

[54] CIDOC-CRM, Modelling Intellectual Processes: the object-orient FRBR model, http://cidoc.ics.forth.gr/docs/frbr_oo/FRBR_tutorial_gothenburg.ppt, (accessed 14 May 2007)
[55] Tutorials about RDF and other web related technologies are available at: W3 Schools, http://www.w3schools.com/default.asp, (accessed 15 May 2007)
[56] W3 Schools, RDFS, http://www.w3schools.com/rdf/rdf_schema.asp, (accessed 15 May 2007)

encoding can assist with the handling of data. *METS*[57] (Metadata Encoding and Transmission Standard) is a schema for descriptive, administrative and structural metadata about objects within a digital library and requires file inventory information to be included (listing all files associated with the digital object) and a structural map which outlines the hierarchical structure linking the object with its content files and metadata. *EAD*[58] (Encoded Archival Description) is based on the *TEI* Guidelines[59] (Text Encoding Initiative) and is aimed at library and museum professionals who need to make finding-aids such as inventories, indexes and registries machine-readable. The enthusiastic global adoption of *TEI* across multiple academic disciplines makes this encoding method an important element in strategies to do with information management. Not only are the *TEI* Guidelines a standard way of expressing information about textual data, but they have also been extended for use in more specific realms, e.g. *CES*[60] (Corpus Encoding Standard – and its XML instantiation, XCES) for describing corpora; and *MEP*[61] (Model Editions Partnership) for creating editions of historical documents.

Information Retrieval and Analysis

The term Information Retrieval (IR) covers a multitude of processes and techniques but is often used in contexts where there is significant input from 'computing' (rather than 'information') science. Eidenburger presents evidence (based on papers relating to image and video retrieval deposited with the IEEE[62] digital library) that suggests that the numbers of papers submitted to scientific journals that address IR research has been steadily increasing over the last twenty years.[63] The field also has a number of conferences that are dedicated to discussing various areas of IR research e.g. ECIR[64] (European Conference on Information Retrieval), ISMIR[65] (International Conference on Music Retrieval), CIVR[66] (ACM International Conference on Image and Video Retrieval), so it is clear that research into this area is well funded and offers attractive challenges to computer science.

Corporate scale search engines such as *Google* and *Yahoo* inevitably dominate the wider context of the information retrieval landscape and it is fairly clear that for many users, these tools are always their first (and sometimes their last!) resort. There are valid concerns that where some kind of authorisation or an acknowledgement of terms and conditions is required to access data, web crawlers are unable to index this material, some of which is held within HE sector institutional digital repositories, resulting in a great mass of data that is 'hidden' from general Internet browsing. With the advent of *OAI-PMH* compliant repositories however, the major search engines are increasingly able to include these 'deep web' resources into their search results. Whereas other search engines rely on *OAI-PMH* repositories choosing to register, *Yahoo* has set up an agreement with *OAISter*,[67] a search engine targeted specifically at OAI repositories, to share access to a wide range of academic material hosted by libraries and institutions all over the world.[68] Consequently, a study carried out in 2006[69] reported that of the 3.3 million unique web resources described

[57] Library of Congress, Metadata Encoding and Transmission Standard, http://www.loc.gov/standards/mets/, (accessed 15 May 2007)
[58] Library of Congress, Encoded Archival Description, http://www.loc.gov/ead/eaddev.html, (accessed 15 May 2007)
[59] Text Encoding Initiative Consortium, http://www.tei-c.org/Guidelines2/, (accessed 15 May 2007)
[60] Vassar College, Corpus Encoding Standard, http://www.cs.vassar.edu/CES/, (accessed 15 May 2007)
[61] University of Illinois at Chicago, Model Editions Partnership, http://tigger.uic.edu/~cmsmcq/mep/mepw04.html, (accessed 15 May 2007)
[62] Institute of Electrical and Electronics Engineers
[63] See Eidenberger (2004), see also, Methods Network, What's in the Art Historian's Toolkit?, Working Paper, http://www.methodsnetwork.ac.uk/redist/pdf/wkp1.pdf, (accessed 15 May 2007)
[64] ECIR, http://ecir2006.soi.city.ac.uk/, (accessed 16 May 2007)
[65] ISMIR, http://ismir2006.ismir.net/, (accessed 16 May 2007)
[66] CIVR, http://www.civr2007.com/, (accessed 16 May 2007)
[67] OAIster, http://www.oaister.org/, (accessed 16 May 2007)
[68] University of Michigan, News Service, March 2004, http://www.umich.edu/news/index.html?Releases/2004/Mar04/r031004, (accessed 16 May 2007)
[69] McCown et al (2006)

in the *Dublin Core* metadata available from *OAI-PMH* repositories, the following percentage of material was discovered using three different search engines:

- 65% - Yahoo
- 44 % - Google
- 7% - MSN
- 21% - Material not indexed by the above search engines

Integration of the retrieval power of major search engine technology has been a feature of some academic projects, an example of which is *Syllabus Finder,* developed by Dan Cohen at the CHNM[70] (Centre for History and New Media). Keywords relating to relevant syllabi are identified by word frequency analysis on a set of documents that are known to be relevant to the topic. These keywords are then bundled up with the desired search term and are used in conjunction with the Google Custom Search Engine,[71] which then delivers highly relevant results.[72] An augmented or intelligent search process such as this might begin to define the broad area of research known as 'data mining' (or 'text mining').

Witten at al[73] describe the use of the *GATE*[74] (General Architecture for Text Engineering) development environment in combination with the *Greenstone Digital Library System* and conclude that the added text mining functionality that *GATE* provides can be very successfully embedded. *Greenstone* features a modular architecture which accommodates plug-ins for various functions and this includes a number text mining subsystems. One example includes a process to extract acronymns and their definitions from the full text of a collection. Another process extracts key phrases contained in the text of documents and adds them as metadata. A third 'computes a hierarchy of all phrases contained in the text of the documents and allows the user to browse it, optionally in conjunction with a standard thesaurus'.[75] *GATE* however provides richer functionality incorporating many tasks associated with linguistics research, e.g. part-of-speech tagging, tokenization, sentence splitting, all functions which can assist the semantic tagger, a central feature of the system. It also provides access to resources such as lexicons and ontologies and comes with its own lightweight information extraction system called *ANNIE*, a component which, amongst other tasks, is able to detect person and organization names, geographical locations, dates, times and money amounts. Witten et al conclude that by customizing a digital library with enhanced text mining capabilities, users can experience the type of advantages that the 'semantic web' promises.

The U.S. based project *NORA* Project applies data mining techniques to a collection of literary texts and uses the *D2K* framework developed at the University of Illinois' National Center for Supercomputing Applications (NCSA).[76]

> Data to Knowledge is a rapid, flexible data mining and machine learning system that integrates analytical data mining methods for prediction, discovery, and deviation detection, with data and information visualization tools.[77]

Fuller references to the architecture of this system can found elsewhere[78] but it may be useful to highlight the visualization feature that is a component part of this system (see fig.5). This paper, for reasons of

[70] George Madison University, CHNM, http://chnm.gmu.edu/, (accessed 16 May 2007)
[71] See also, OpenDAOR (Directory of Open Access Repositories) and their use of the Google engine, http://www.opendoar.org/search.php, (accessed 24 May 2007)
[72] For further reference, see, Methods Network, Tools and Methods for Historical Research, Working Paper, http://www.methodsnetwork.ac.uk/redist/pdf/wkp4.pdf, (accessed 16 May 2007)
[73] Witten et al (2004)
[74] University of Sheffield, Natural Language Processing Group, GATE, http://gate.ac.uk/, (accessed 16 May 2007)
[75] Witten et al (2004), p.2
[76] NCSA, Automated Learning Group, http://alg.ncsa.uiuc.edu/do/index, (accessed 16 May 2007)
[77] NCSA, Automated Learning Group, D2K: Data to Knowledge, http://alg.ncsa.uiuc.edu/do/tools/d2k, (accessed 16 May 2007)

brevity and to avoid encroaching on territory covered by other working papers, has largely conceptualized the 'library' as a collection of texts that can be managed, manipulated and visualized using text-specific tools. *NORA* and a host of other tools[79] offer users alternative ways of visualizing the results of data analysis processes.

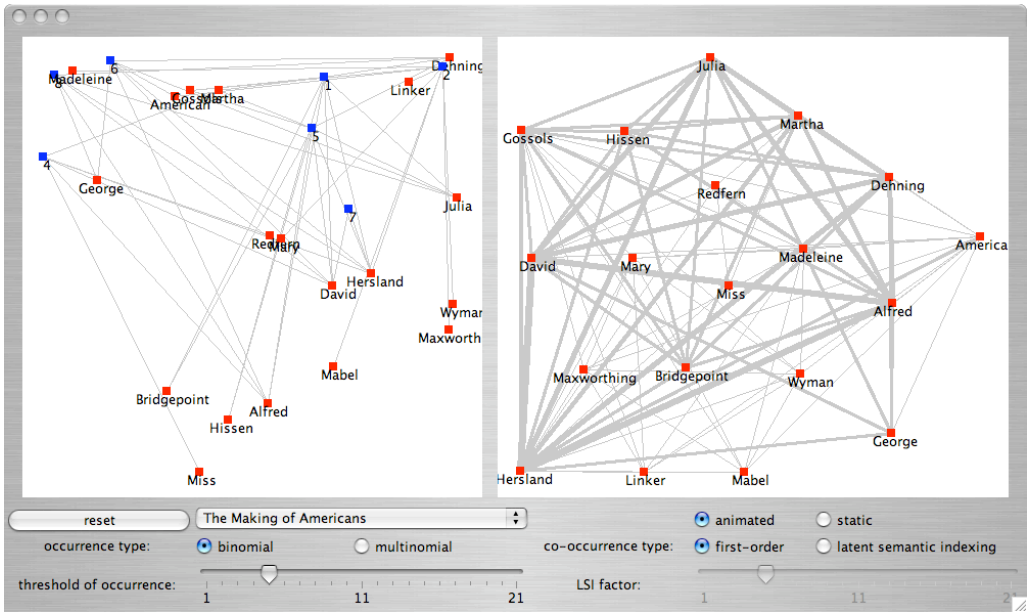

Fig.5 Screenshot from *NORA* social network demo[80]

User Issues

The mass adoption of Web 2.0 type resources such as *FlickR*, *Myspace*, *YouTube* and *Del.icio.us* means that large numbers of users are now attempting to archive and describe digital objects without recourse to expert classification systems. This is widely acknowledged to have positive and negative implications in terms of the validity and effectiveness of data,[81] but in terms of systems design, there is an interesting 'desire line' principle that might be applicable to the construction of indexes and thesauri. 'Desire lines' are the worn areas of a landscape, usually quite narrow tracks, where walkers have decided to take a route that doesn't correspond to, or is completely at odds with, a formal path. Peter Merholz suggests that 'a smart landscape designer will let wanderers create paths through use, and then pave the emerging walkways, ensuring optimal utility.'[82]

Taking this approach as a model for building effective vocabularies might mean that they take a long time to become established, but used in combination with existing systems, this method could add value. At a recent Methods Network workshop,[83] one of the participants mentioned that the British Library records all search terms entered into the catalogue – whether they return successful results or not – in order to ascertain what search terms people would prefer to use, rather than simply being able to gauge which

---

[78] For further reference, see, Methods Network, Tools and Methods for Historical Research, Working Paper, http://www.methodsnetwork.ac.uk/redist/pdf/wkp4.pdf, (accessed 16 May 2007)
[79] see: Transliteracies Project, http://transliteracies.english.ucsb.edu/category/research-project/research-clearinghouse-individual/objects-for-study-individual/text-vizualization/, (accessed 16 May 2007)
[80] Nora Project, http://www.noraproject.org/social_network.php, (accessed 16 May 2007)
[81] For discussion of the pros and cons of folksonomies, see: Mathes, (2004)
[82] Adaptive Path, Metadata for the Masses, http://www.adaptivepath.com/publications/essays/archives/000361.php, (accessed 17 May 2007)
[83] Methods Network, The Future of Information Technology in Music Research and Practice, Goldsmith's College, 8 September 2006, http://www.methodsnetwork.ac.uk/activities/wsp6.html

terms from an existing list are used most regularly. Clearly there is an argument for saying that a lot of community-generated terminology will be faddish and that terms that appear to have significant collocation at one point in time (e.g. Iraq/war-torn – circa 2007) may be subsequently misleading. Counter arguments might point to the value of such conceptual relationships as chronological markers, if temporal data can be applied to the linkages between terms.

The more orthodox approach in terms of LIS principles is to provide structured, secure and authoritative systems that ensure appropriate users have access to licensed and reliable sources of information. One of the enabling components of this strategy is to be able to authenticate and authorise users to access relevant systems and a key tool that is currently being promoted by JISC and the UK Federation for Access Management[84] is called *Shibboleth*.[85] It defines a set of protocols written in *SAML* (Security Assertion Markup Language) that enable the passing of secure identity information between institutions and service providers and it allows organisations to provide a single logon for users to access all of the local and remote resources that they are authorised to use. The principle advantage to the user is that they are no longer required to remember a multitude of passwords to access a range of distributed resources. The institution benefits by no longer being reliant on third parties to reset accounts when problems arise, as the devolved architecture of *Shibboleth* means that the process of establishing *who you are* (authentication) is separate from the process of establishing *where you can go* (authorisation), the former being entirely dependent on a single locally-assigned password.

Getting the balance right between creating secure and robust systems, whilst at the same time keeping them attractive, intuitive and usable, is a significant area of LIS research. The use of online surveys and the collection of statistics via transaction log analysis (TLA) are two ways of building up a picture of how users engage with web resources, although both methods have limitations as well as strengths. Online surveys can be valuable for revealing user's motivations, goals, attitudes and satisfaction levels but it is very difficult to ensure that the survey sample is representative of the user group that is being targeted. Another problem is that users are increasingly reluctant to engage with questionnaires,[86] a factor which does not impact upon TLA methods as data is automatically harvested from computerized log files and provides the analyst with an objective view of user behaviour rather than having to rely on error-prone human accounts of activity. Where TLA methods fall short however, is their inability to accurately identify and track individual users of systems. Log files rely on IP addresses or hostnames to identify client computers and there is no way of ensuring that use of the client system is limited to one defined user. Cached pages and unattended active sessions may also cause problems with the reliability of the results. A combination of both approaches however can prove to be very effective for understanding website usage.

COUNTER (Counting Online Usage of Networked Electronic Resources) is an international initiative that focuses on the effective recording and exchange of online usage statistics. Aimed at libraries, publishers and software vendors, it also liaises with a number of other organisations, one of which, the Association of Research Libraries (ARL), is responsible for a suite of tools and services called *StatsQUAL*.[87] Comprising of three different component systems, *LibQUAL+*, *DigiQUAL* and *MINES for Libraries,* they are designed to enable effective assessment of the role, character and impact of physical and digital libraries.

One of the critical tasks of any systems development project is to ensure that the human-computer interface (HCI) is thoroughly tested and is entirely fit for purpose. Zabed et al[88] give details of a sophisticated iterative process of designing and evaluating an IR interface and recount the eight-step

[84] UK Access Management Federation for Education and Research, http://www.ukfederation.org.uk/, (accessed 17 May 2007)
[85] Internet 2, Shibboleth project, http://shibboleth.internet2.edu/, (accessed 17 May 2007)
[86] Evidence suggests that response rates for online surveys are lower than for other media and continue to decline. Cited in: Harley and Henke (2007), p.4
[87] Association of Research Libraries Statistics and Measurements Dept, StatsQUAL, http://www.libqual.org/, (accessed 17 May 2007)
[88] Zabed Ahmed et al (2006)

sequence that resulted in the development of a prototype. Offering quantitative statistics over a number of categories relating to relative satisfaction levels of novice and experienced users, the authors manage to arrive at a number of recommendations for IR interface development.

Conclusion

This paper refers to a handful of issues that relate to the use of tools (and their related methods) for LIS research and will hopefully form the basis of an ongoing discussion involving LIS practitioners, some of whom will almost certainly wish to introduce entirely different accounts of the use of tools within the discipline. Judging by the prodigious number of journals that relate to the field,[89] the LIS community is an active and collaborative group of professionals and one which has an enormous amount to offer other arts and humanities disciplines in terms of expertise and the exchange of knowledge. It is clear that where certain disciplines have been able to partially ignore or only reluctantly embrace all of the digital systems and methods that have appeared over the last twenty five years, LIS professional have had no choice but to be in the vanguard, dependent as all other researchers are on being able to effectively find resources and continue their research irrespective of publishing formats or system compatibility issues. As one commentator puts it, 'New technologies have created new communities. It is the task of libraries to invent new information services'.[90]

Neil Grindley
Methods Network
Senior Project Officer
Original Draft – April 2007
Version control – 05 July 2007

References

Besser, H. (2004), 'The Past, Present, and Future of Digital Libraries', in Schreibman, S., Siemens, R., Unsworth, J., (eds), *A Companion to Digital Humanities*, (pp.557 - 575)

Biglan, A. (1973), 'The Characteristics of Subject Matter in Different Academic Areas', *Journal of Applied Psychology*, 57, pp. 195-203

Eidenberger, H. (2004), *A New Perspective on Visual Information Retrieval*, Vienna University of Technology, http://www.ims.tuwien.ac.at/~hme/papers/ei2004-vir.pdf, (accessed 17 May 2007)

Harley, D., Henke, J. (2007), Toward an Effective Understanding of Website Users: Advantages and Pitfalls of Linking Transaction Log Analyses and Online Surveys, D-Lib Magazine, March/April 2007, Volume 13, Number ¾ , http://www.dlib.org/dlib/march07/harley/03harley.html (accessed 17 May 2007)

Hey, J. (2004), The Data, Information, Knowledge, Wisdom Chain: The Metaphorical Link, Publisher: Intergovernmental Oceanographic Commission (UNESCO), see: http://ioc.unesco.org/Oceanteacher/OceanTeacher2/02_InfTchSciCmm/DIKWchain.pdf

Lewis, E. (2001), The Role and Impact of the Internet on Library and Information Services, Greenwood Press, Westport Connecticut,

[89] The University of Sheffield Library subscription list of Library and information Science e-journals is indicative of the number available, see: University of Sheffield, E-journals Library & Information Science, http://www.shef.ac.uk/library/elecjnls/ejlib.html, (accessed 18 May 2007)
[90] Lewis (2001)

McCown, F., Liu, X, Nelson, M., et al (2006), 'Search engine Coverage of the OAI-PMH Corpus', *IEEE Computer Society*, March/April 2006 (Vol. 10, No. 2) pp. 66-73, http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/mags/ic/&toc=comp/mags/ic/2006/02/w2toc.xml&DOI=10.1109/MIC.2006.41, (accessed 17 May 2007)

Mathes, A. (2004), Folksonomies – Cooperative Classification and Communication Through Shared Metadata, http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html, (accessed 17 May 2007)

Norman, D. A. (1999), *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is so Complex, and Information Appliances Are the Solution*. Boston: MIT Press

Shera, J.H. (1983), 'Librarianship and Information Science', in Machlup, F., Mansfield, U. (eds), *The Study of Information: Interdisciplinary Messages*, Wiley, New York, pp. 379-388

Webber, S. (2003), 'Information Science in 2003: a critique', *Journal of Information Science*, 29 (4), pp. 311-330

Witten, I., Don, K, Dewsnip, M., et al (2004), 'Text Mining in a Digital Library', *International Journal on Digital Libraries*, Volume 4, Number 1, August 2004, pp. 56-59(4), see also: http://www.dcs.shef.ac.uk/~valyt/download/greenstone-gate.pdf

Zabed Ahmed, S., McKnight, C., Oppenheim, C., 'A User-Centred Design and Evaluation of IR interfaces', *Journal of Librarianship and Information Science*, 38 (3) September, 38 (3), pp.157-172, http://lis.sagepub.com/cgi/reprint/38/3/157.pdf, (accessed 17 May 2007)

LINKS

The following are a sample selection of links that indicate the range of material that relates to the Digital Library and Information Studies sector. Some of these duplicate footnote references and all were consulted in the course of researching this paper. They are in no particular order within content headings

Contents

Articles and Reports
Journals
Ontologies and Semantic Web
Organizations and Projects
Repositories and Digital Libraries
Resources and Listings
Standards
Tools
Web

Articles and Reports

Internet Librarian
http://www.internet-librarian.com/2005Presentations/
Conference proceedings

Review in D-Lib
http://www.dlib.org/dlib/april02/04bookreview.html
Marilyn Deegan and SimonTanner – Digital Futures series

EC Open Access Policy
http://openaccess.eprints.org/index.php?/archives/198-guid.html
Lobby to provide open access to research results
Report of NSF workshop on Research Directions for Digital Libraries
http://www.sis.pitt.edu/%7Edlwkshop/report.pdf
Knowledge Lost in Information

DLib – OAIS and OAI
http://www.dlib.org/dlib/april01/04editorial.html
short article outlining difference in terms

5 step guide to being a JISC content provider
http://www.ariadne.ac.uk/issue33/info-environment/
guide to data management using JISC guidelines

Search Engine Coverage of the OAI-PMH Corpus
http://library.lanl.gov/cgi-bin/getfile?LA-UR-05-9158.pdf
article on various search engines and their coverage of OAI-PMH archives

Journals

Journal of Information Science
http://jis.sagepub.com/

Journal of Librarianship and Information Science.
http://lis.sagepub.com/

Journal of the American Society for Information Science and Technology
http://www.asis.org/Publications/JASIS/jasis.html

Annual Review of Information Science and Technology http://www.asis.org/Publications/ARIST/

Sheffield List of e-journals for Lib & Info Sci
http://www.shef.ac.uk/library/elecjnls/ejlib.html

B2i search mechanism for lib and info science material
http://www.b2i.de/index.mhtml?language=eng
German site gives access to journal articles

Ontologies and Semantic Web

Ontologies and Digital Libraries
http://dcs.vein.hu/CIR/cikkek/MFIR_DLOntology4.pdf
Towards a Digital Library Theory: A Formal Digital Library Ontology

OWL – Web Ontology Language
http://www.w3.org/TR/owl-ref/
A semantic markup language for publishing and sharing ontologies on the World Wide Web
Resource Description framework
http://www.w3.org/TR/2004/REC-rdf-primer-20040210/
Primer on the RDF

Metamodel
http://www.metamodel.com/article.php?story=20030115211223271
Community site for meta-modeling and semantic modeling

Jena
http://jena.sourceforge.net/
Semantic web framework for Java

Performing Arts Conceptual Model
http://cidoc.ics.forth.gr/docs/2006_LeBoeuf_eng.pdf
Patrick LeBeouf on conceptual models and definition of the semantic web

Organizations and Projects

Digital Preservation Coalition
http://www.dpconline.org/graphics/join/projects.html
To foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.

Google Directory – Digital Library Development
http://directory.google.com/Top/Reference/Libraries/Library_and_Information_Science/Digital_Library_Development/
List of major organisations involved with digital libraries and curation

H-Info
http://www.h-net.org/~info/
H-Info focuses on the interdisciplinary, international study of information and information institutions, broadly construed

Research Libraries Group
http://www.rlg.org/
RLG Programs is part of OCLC's Programs and Research division. OCLC is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing information costs.

Library and Information Research
http://www.lirg.org.uk/lir/directory.htm
Directory of current research in LIS departments in UK and Ireland

Good Management Practice for Higher Education
http://www.gmp.ac.uk/projects/GMP313
List of projects related to management practices in HE
Research Support Libraries Group
http://www.rslg.ac.uk/research/
Research page

University College Dublin
http://www.ucd.ie/sils/research.htm
School of Information and Library Studies

Strathclyde Centre for Digital Library Research
http://cdlr.strath.ac.uk/
Centre for research based activities

HILT
http://hilt.cdlr.strath.ac.uk/
High Level Thesaurus Project

UK Federation
http://www.ukfederation.org.uk/
UK Access Management federation – shibboleth

EGEE
http://www.eu-egee.org/
EU initiative for Grid computing and sciences

Sheffield Dept of Information Science
http://www.shef.ac.uk/is/research/groups
Research Groups

Repositories and Digital Libraries

Australian Repositories Online for the World
http://search.arrow.edu.au/apps/ArrowUI/
Initiative to manage online resources

Virginia Tech Digital Library and Archives
http://scholar.lib.vt.edu/theses/
Electronic Theses and Dissertations

Architecture of Digital Libraries
http://discovery.bits-pilani.ac.in/rahul/I-Net/Complete-InetBook-PHI-2003-Secure.pdf
Rahul Pilani on the various architectures behind Digital Libraries

Glasgow Digital Library
http://gdl.cdlr.strath.ac.uk/
Organised digital collections to support teaching, learning and research

OpenDoar
http://www.opendoar.org/
*Open*DOAR is an authoritative directory of academic open access repositories – also provides tools and support

JISC digital repositories http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx
lots of links to funded initiatives

DSpace at Loughborough
https://dspace.lboro.ac.uk/dspace/
institutional repository

DLibra
http://dlibra.psnc.pl/
Digital Repository System from the Poznan Supercomputing and Networking Center

EPrints
http://www.eprints.org/software/
Digital Repository platform used by the OU

Open Research Online
http://oro.open.ac.uk/information.html
EPrints based research repository

Fedora Open Source Software
http://www.fedora.info/
Digital Repository Software

Open Repositories Conference
http://openrepositories.org/call/
List of topics relating to Digital Library research and development

AHDS Report on Digital Repository Issues
http://ahds.ac.uk/preservation/curl/curl_case_study_d4.pdf

OAI Interoperability Summit Participants
http://msc.mellon.org/Meetings/Interop/Participants
Powerful group of interests representing repository software and processes

Resources and Listings

Open Directory Project
http://dmoz.org/
Human edited web subject resource

Librarian's Internet Index
http://lii.org/
Another subject resource

BUBL
http://bubl.ac.uk/
Another subject resource tool from Centre for Digital Library Research - Strathclyde

Digital Librarian
http://www.digital-librarian.com/
Librarian's Choice of the best of the web

Training video about UK federation
http://www.jisc.ac.uk/whatwedo/themes/access_management/federation/animation
Introduction to the use of Shibboleth model for standardising access to resources

Evalued
http://www.evalued.uce.ac.uk/
DIY user surveys

OAIster
http://www.oaister.org/
Union catalogue of OAI digital resources

Standards

Digital Library Federation
http://www.diglib.org/standards/draftbmark.htm
Draft benchmark for digital reproductions of printed books and serial publications

Open archives organisation
http://www.openarchives.org/
open archives initiative

OCLC
http://purl.oclc.org/docs/new_purl_summary.html
PURL persistent naming convention

DELOS
http://www.delos.info/index.php?option=com_content&task=view&id=345&Itemid=
Digital Library Reference Model System

Shibboleth
http://shibboleth.internet2.edu/
Shibboleth is standards-based, open source middleware software which provides Web Single SignOn (SSO) across or within organizational boundaries.

Search Tools
http://www.searchtools.com/info/classifiers.html
HEA recommended guide to classification systems

Functional Requirements for Bibliographic Records
http://cidoc.ics.forth.gr/docs/frbr_oo/FRBR_tutorial_gothenburg.ppt
Is a core model to integrate library information

METS
http://xml.coverpages.org/mets.html
Metadata Encoding and Transmission Standard

Tools

Seattle Central Community College
http://www.seattlecentral.org/faculty/jshoop/reftools.html
Reference Tools

Robotics Intelligence Dept, Jaume I University, Castellon, Spain
http://www.formatex.org/micte2005/168.pdf
Learning Visual Servoing Techniques by Remotely Programming an Internet Tele-Lab

Robot Gossip Blog
http://robotgossip.blogspot.com/2005/10/robot-librarian.html
UJI Robot Librarian

RFID at Leeds
http://www.cilip.org.uk/publications/updatemagazine/archive/archive2007/april/loughranandyoung.htm
Case study of RFID implementation

Berekely Digital Library Sunsite
http://sunsite.berkeley.edu/Tools/
Tools for Building the Digital Library

Akamai
http://www.akamai.com/
Content Delivery, application performance management and streaming media services

University of British Columbia
http://www.cs.ubc.ca/rr/research/toolkit.shtml
Librarian's Toolkit

Personal Construct theory
http://www.springerlink.com/content/jhmg4dc2v9w0mqux/
Mary Burke on PCT as a research tool for analyzing user perceptions of photographs

Sakai Briefing
http://www.hull.ac.uk/esig/downloads/sakai_overview_Dec05.pdf
University of Hull Sakai presentation

uPortal
http://www.uportal.org/index.html
Free open-standards based collaboratively developed portal tool

CIDOC-CRM tools
http://cidoc.ics.forth.gr/tools.html
mapping tools for CRM onto XML types

Text Mining, GATE and Greenstone
http://www.dcs.shef.ac.uk/~valyt/download/greenstone-gate.pdf
Text Mining in a Digital Library

Web

BrightPlanet
http://www.brightplanet.com/resources/details/deepweb.html
White Paper on the hidden web

LaGuardia Community College – Library Media Resources Centre
http://www.lagcc.cuny.edu/library/invisibleweb/fastfacts.htm
The invisible Web

Deep Web Research
http://deepwebresearch.blogspot.com/
Very full list of deep web links