World Digital Libraries

an international journal

Volume 2, 2009

The Energy and Resources Institute

Contents

WDL (World Digital Libraries)
Volume 2, Number 1 • June 2009

Editorial

N Balakrishnan
Indian Institute of Science, Bangalore

World Digital Libraries 2(1) v–vi

It is very common now for scientists across multiple disciplines to access and mine data and information from a variety of heterogeneous sources. The information varies from textual to numeric, medical to geospatial, and quantitative to qualitative. This is stimulated by the fact that future innovation lies in the interface of multiple disciplines, particularly between engineering and science, and the trend towards borderless research. The borders between disciplines and between geographic locations have become truly porous.

In such a world, the efforts of information scientists are directed towards innovating new tools and techniques to convert such a vast resource of data and information to knowledge that can lead to innovation, thus heralding an era of exponential growth of knowledge. One such direction is the 'Semantic Web'. Though the concept of the Semantic Web has been around for several years, its importance and applications to newer domains was not recognized in the past. In this issue, Kim H Veltman has looked at the present and the future of the Semantic Web. His gaze into the crystal ball brings out the future that will be in tune with the integration of the 'Internet of Things' and the need to integrate ideas to distinguish and search knowledge from different worlds- metaphysical, man-made, social, and creative.

In the world dominated by heterogeneity in types of information, services, and tools, interoperability is a daunting need. The paper by Dennis Nicholson, Anu Joseph, and Emma McCulloch describes a novel idea of subject interoperability through embedding pilot terminology web services. The authors present an expert evaluation of DiLAS (Digital Library Annotation Service). Their work, as expected, has culminated in the generation of large number of user requirements for the next prototype.

Digital technology, while offering all the comforts of providing instant and democratic access to information, is also known to be fragile. Fast obsolescence is very common. It is indeed a great challenge today to preserve the digital knowledge over a period of time. Heather Brown's paper is one with high practical utility and describes three key strategies that may potentially have the capacity to profoundly affect people's understanding of digital preservation.

One of the important roles of the World Digital Libraries Journal is to share international experiences. We are very fortunate to have two articles - one by Jakob Heide Petersen and Jens Thorhauge on Danish digital libraries strategy and the other by Kristiina Hormia-Poutanen on digital library infrastructure in Finland. Both these articles describe some of the important issues that one

needs to address in the world of digital libraries and also a few of their current projects.

We hope that you will enjoy reading these and would benefit from them. We also hope that these articles, through which world renowned experts share their experiences, would stimulate many of you to contribute and work towards the advancement of digital libraries.

HILT IV: subject interoperability through building and embedding pilot terminology Web services

Dennis Nicholson,* Anu Joseph, and Emma McCulloch

World Digital Libraries 2(1): 1–22

Abstract

A report of work carried out within the JISC-funded HILT Phase IV project, the paper looks at the project's context against the background of other recent and ongoing terminologies work, describes its outcome and conclusions, including technical outcomes and terminological characteristics, and considers possible future research and development directions. The Phase IV project has taken HILT to the point where the launch of an operational support service in the area of subject interoperability is a feasible option and where both investigation of specific needs in this area and practical collaborative work are sensible and feasible next steps. Moving forward requires detailed work, not only on terminology interoperability and associated service delivery issues, but also on service and end user needs and engagement, service sustainability issues, and the practicalities of interworking with other terminology services and projects in UK, Europe, and global contexts.

* Director, Centre for Digital Library Research, Information Resources Directorate, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow, G1 1XH
E-mail d.m.nicholson@strath.ac.uk

Introduction and overview

This paper reports on the work carried out within Phase IV of the JISC-funded HILT[1] project, considers its context against the background of other recent and ongoing terminologies, describes its outcome and conclusions, and considers future research and development directions, some of which are envisaged as being collaborative. As with HILT Phase III (Nicholson and McCulloch, 2006a; 2006b), the project utilized the expertise of participants at CDLR[2] and EDINA,[3] and of the HILT terminology advisors,[4] together with some ongoing liaison with MIMAS,[5] who run IESR[6] and Intute.[7] It also had ongoing support from OCLC[8] with respect to allowing the use of the electronic files of DDC[9] (Dewey Decimal Classification) and of LCSH[10] mappings to DDC[11] and of a range of other participants on the HILT Steering group.[12] The project is working towards setting up a JISC Shared Infrastructure Service with respect to terminologies and interpretability. Shared Infrastructure Services are an important building block of the JISC IE (Information Environment) architecture.[13] They constitute a common set of infrastructural services that provide this information through M2M (machine-to-machine) interfaces, on which other elements of the environment, such as portals, brokers and aggregators, can draw. In the case of HILT, the focus is on the provision of M2M access to terminology and terminology interoperability data for other services to interact with and use.

Ensuring that HE (higher education) and FE (further education) users of the JISC IE can find appropriate learning, research and information resources by subject search and browse in an environment where most national and institutional service providers – usually for very good 'local' reasons – use different subject schemes to describe their resources is a major challenge facing the JISC domain (and, indeed, other domains beyond JISC). Encouraging the use of standard terminologies in some services (institutional repositories, for example) is a related challenge. Under the auspices of the HILT project, JISC has been investigating mechanisms to assist the community with this problem, and thereby, optimize the value obtained from expenditure on content and services by facilitating resource sharing to benefit users in the learning and research communities. The project has been through a number of phases, with work from earlier phases reported, both in the form of published work,[14] and project reports (see the project website). The focus of this paper is on Phase IV, which began in May 2007 and is due to complete a few months from the time of writing at the end of May 2009.

HILT Phase IV had two elements: the core project, the focus of which was 'to research, investigate and develop pilot solutions for problems pertaining to cross-searching multi-subject scheme information environments, as

[1] http://hilt.cdlr.strath.ac.uk/
[2] Centre for Digital Library Research (CDLR): http://cdlr.strath.ac.uk/
[3] EDINA: http://edina.ac.uk/
[4] Alan Gilchrist and Leonard Will
[5] Manchester Information & Associated Services (MIMAS): http://www.mimas.ac.uk/
[6] Information Environment Services Registry (IESR): http://iesr.ac.uk/
[7] intute: http://www.intute.ac.uk/
[8] OCLC Online Computer Library Center: http://www.oclc.org/
[9] Dewey Decimal Classification (DDC): http://www.oclc.org/dewey/
[10] Library of Congress Subject Headings (LCSH): http://authorities.loc.gov/
[11] LCSH to DDC mappings: http://www.oclc.org/asiapacific/zhcn/dewey/updates/numbers/default.htm
[12] http://hilt.cdlr.strath.ac.uk/hilt4/groups.html
[13] http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/
[14] See http://hilt.cdlr.strath.ac.uk/dissemination.html for a list key HILT publications

well as providing a variety of other terminological searching aids, and a short extension to encompass the pilot embedding of routines to interact with HILT M2M services in the user interfaces of various information services serving the JISC community. Only the outcomes of the core HILT IV work are reported here. These are summarized in section HILT IV: Project outcomes, which follows a short section on related work that sets HILT in the context of wider work in the area. The next section describes the current position in the project with respect to available working facilities and project perspectives, adding detail to the section on summary outcomes, and the following section covers the aims of the embedding extension project and addresses future research and development directions beyond the end of HILT IV. The last section provides a brief conclusion to the paper.

Related work

There is a good deal of ongoing and recent work in the general area of terminologies, terminological interoperability in cross-searching and other scenarios, and terminology and terminology interoperability services (Day, Koch, and Neuroth 2004; Landry 2004; Tudhope, Koch, and Heery 2006; Zeng and Chan 2006; van Gendt, *et al.* 2006; Vizine-Goetz, Diane, Houghton, *et al.* 2006; Macgregor and McCulloch 2006: Agosti, *et al.* 2007; Mayr and Petras 2007; Binding, Tudhope, May, 2008; Geser 2008; Levergood, *et al.* 2008). Much of it this of relevance, or potential relevance, to HILT because HILT's intended role as a JISC shared infrastructure service aiming to support the work of end-user-focused information services facing a myriad of different problem situations and their various facets means its aim has largely been to address the interoperability problem in general, rather than some specific aspect of it, since its primary focus is serving researchers, teachers, and learners in universities, and since this community as a whole is likely to need to find and access subject information located almost anywhere in the world and to do so using almost any of the available KOS[15] (knowledge organisation systems), and, potentially, any language, it is impossible to envisage a future service being a stand-alone enterprise. The need for HILT, and for any similar general service, will be to provide, not just direct access to the KOS and KOS interoperability data it provides, but also to offer services using its facilities to permit discovery of, and integrated interaction with, other sources of such data, as well as means of intelligently and transparently handling this M2M data in user interfaces to best meet the needs of individual user groups tackling particular types of task.

This means the project impacts a wide variety of work in the field. The work of significant or recent projects in the area (CACAO,[16] MACS,[17] LIMBER, RENARDUS,[18] KoMoHe,[19] and STAR[20] are examples) is important not just because the issues they face are similar, but because in future, comprehensiveness, and hence the need to interoperate with initiatives of this kind (Nicholson 2008) will be an issue, and their work helps highlight some of the requirements of that. So the work on different approaches to providing cross-works or cross-walk services (Zeng and Chan 2004), whether it be on differing approaches to intellectual

[15] Taxonomies, classification schemes, thesauri, subject heading lists and similar – see Zeng and Chan, 2004 for a full description

[16] http://www.cacaoproject.eu/

[17] MACS Project: http://macs.vub.ac.be/pub/

[18] RENARDUS Project : http://renardus.sub.uni-goettingen.de/

[19] KoMoHe Project http://www.gesis.org/en/research/information_technology/komohe.htm

[20] STAR Project http://hypermedia.research.glam.ac.uk/kos/star/

mapping, on machine processing or automation-based approaches (see for example, Doerr 2001; Chan and Zeng 2002; Koch and Vizine-Goetz 1998; Godby, *et al.* 1999; Ardo 2004) or on the advantages and disadvantages of various different strategies or approaches. For historical (Nicholson and McCulloch 2006a), and to some extent, resourcing reasons, HILT's approach has been based on intellectual mapping, but this does not mean it is seen as the only viable approach in every circumstance. For obvious reasons, work on the advantages of cross-walk work (Mayr and Petras 2008) and on problems encountered (Liang and Sini 2006, McCulloch and Macgregor 2008) is of direct benefit, whether for current work or future planning. Multi-lingual efforts in the area (Cacao 2008; Clavel-Merrin 2004; Landry 2004; Agosti, *et al.* 2007) are significant because various groups in the academic community that is the project's focus have research and teaching and learning interests that cross language boundaries. Since there is an assumption that a future service will have to integrate other (perhaps unknown) terminology services into its operations for the benefit of those it serves, and provide mechanisms to aid the discovery of simple information services covering particular subject areas, work on terminologies and services registries is of interest (Apps 2007; Hillman, *et al.* 2006; Golub and Tudhope 2008). And since the aim is to aid information services in the provision of enhanced (and transparent) subject retrieval and browse facilities for their users and staff via M2M interactions with HILT, and the subsequent intelligent and user-adaptive handling within their user interfaces of returned terminological data encompassing different KOS and KOS cross-walk approaches with differing characteristics, work on user interfaces (Shiri 2008; Stafford, *et al.* 2008; Shiri, Nicholson, and McCulloch 2004), work on user behaviour and what motivates it (Blair 1980; Brown 1995; Buckland, Chen, Gebie, *et al.* 1999; McCray, Loane, Browne, *et al.* 1999; McCulloch, Shiri, and Nicholson, *et al.* 2004; Teevan, Alvarado, Ackerman, *et al.* 2004), and similar, and KOS typologies (Tudhope 2006) is also relevant (for example, when designing the embryonic toolkit described below).

It would not be difficult to add to this list of related work, but it is unnecessary here. A flavour of the complexity of possible relationships is all that is required to provide context and a feel for the landscape of potential interconnections. A detailed investigation of the relevance and implications of the work referenced above for future services of the kind envisaged by HILT would be relevant and illuminating, but is neither appropriate nor feasible in the present paper, which is concerned primarily with the work of HILT IV itself. What is perhaps worth mentioning is that the one thing missing from the plethora of ongoing work in the area from the HILT perspective (and, arguably, that of similar enterprises such as the OCLC terminology services suite[21]) is co-ordination of R&D work within an agreed common architecture or architecture set, an issue covered further in the subsequent section.

HILT IV: project outcomes

Prior to its extension to encompass a short project to embed the pilot use of HILT M2M services in a variety of service interfaces, HILT IV was scheduled to complete in February 2009. By March 2009, therefore, this core work was complete. The project had

- Built and tested a range pilot M2M (SRU/W,[22] SOAP,[23] and SKOS[24]) based web

[21] http://www.oclc.org/research/projects/termservices/resources/termservices-overview.pdf
[22] Search/Retrieve Web Service (SRU/W): http://www.loc.gov/standards/sru/
[23] SOAP: http://www.w3.org/TR/soap/
[24] Simple Knowledge Organization System (SKOS) Core: http://www.w3.org/2004/02/skos/

services to deliver terminologies and terminology mappings to JISC and institutional information services, supporting the transparent enhancement of subject search facilities.
- Built and tested a database of terminologies (DDC,[25] UNESCO,[26] HASSET,[27] IPSV,[28] JITA,[29] MeSH,[30] CAB, GCMD,[31] NMR,[32] AAT,[33] SCAS, JACS,[34]), high level mappings to DDC (HASSET, IPSV, UNESCO, JACS), and limited sample 'deeper' mappings to MeSH and HASSET.[35]
- Built and tested an embedding toolkit to offer information services the core software needed to begin building subject interoperability services, for their users by interacting at M2M level with the above- mentioned web services and database through routines embedded transparently in their service user interfaces—a programmer's toolkit to help build improved subject browse and retrieve facilities.
- Conducted various practical experiments to successfully embed terminology service interaction into JISC community services to create operational pilot subject browse, retrieve, and deposit enhancements for service users (a smaller piece of work preceding the embedding extension project mentioned before).
- Developed a generic terminology services architecture that will (over a period of time) permit the HILT services to grow and improve by incorporating terminology services being developed elsewhere.
- Determined that a terminology services registry is a key part of this architecture and that the core functionality required to build and run such a registry is already inherent in HILT pilot services.
- Developed staff skill sets and experience associated with the problems of subject searching and subject interoperability within and across information services using different subject terminologies, with the best approaches to mapping new schemes into the database, and with an associated distributed architecture to permit the ready integration of new services into the JISC and global subject interoperability landscape.
- Determined in conjunction with JISC that the best general option for a sustainable operational Shared Infrastructure Service based on HILT project outcomes is a Terminologies Interoperability Centre offering a mix of standard 'plug and play' type M2M and toolkit facilities free at the point of use, including a training portal and an associated terminology services registry, more flexible, charged-for, special versions of these, tailored to the needs of individual services and institutions, and ongoing development via a mix of collaboration and externally funded R&D, as well as through JISC support.

In consequence, it had developed to the point where it was in a position to tackle the more advanced embedding work required by the extension project (described in outline in

[25] Dewey Decimal Classification (DDC): http://www.oclc.org/dewey/
[26] UNESCO Thesaurus: http://www2.ulcc.ac.uk/unesco/
[27] Humanities and Social Science Electronic Thesaurus (HASSET): http://www.data-archive.ac.uk/search/hassetSearch.asp
[28] Integrated Public Sector Vocabulary (IPSV): http://www.esd.org.uk/standards/ipsv/
[29] http://eprints.rclis.org/jita/
[30] Medical Subject Headings (MeSH): http://www.nlm.nih.gov/mesh/
[31] Global Change Master Directory (GCMD): http://gcmd.nasa.gov/Resources/valids/keyword_list.html
[32] National Monuments Record Thesauri (NMR): http://thesaurus.english-heritage.org.uk/
[33] Art & Architecture Thesaurus (AAT): http://www.getty.edu/research/conducting_research/vocabularies/aat/
[34] Joint Academic Coding System (JACS): http://www.ucas.ac.uk/figures/ucasdata/subject/
[35] Also included are LCSH mappings included with the DDC file provided by OCLC.

the subsequent section), and put forward to JISC proposals for future development towards setting up of an operational shared infrastructure service. It had in place a core of tested pilot facilities and terminology sets and an embryonic mappings set that could be the basis of a future service, an understanding of the terminological, technical, architectural, and staffing requirements for future development, and a clear if outline vision of what was required to move towards a useful, robust, and sustainable service aligned with JISC's strategic aims.[36]

HILT IV: working facilities and project perspectives—the current position

The current position in the project is best described under four headings: Service functions, database, and embedding toolkit... the HILT architecture and its implications; project understanding of service and end-user needs; and the requirements of a robust, useful, and sustainable service.

Service functions, database, and embedding toolkit

A complete description of these and how these functions in practical circumstances require greater detail than is possible here and is provided in McCulloch (2008). The present paper deals only with the technical details of the working facilities and with possible ways in which they can be used by information services through M2M interaction with HILT.

Technical summary

Although the following description omits some of the detail, it gives a useful overview of the technical basis of the HILT services.

The system is designed in a modular fashion and Figure 1 shows how each of the different modules interact. This design facilitates the replacement or modification of the components, if a need arises in the future. SQL Server 2005 is a highly scalable, programmable, secure database management system which is used for data storage. The SOAP server enables a distributed web based environment and provides the APIs to interact with the database. A SRW/U server is implemented to support M2M interaction and the APIs can be accessed through the SRW/U protocol.

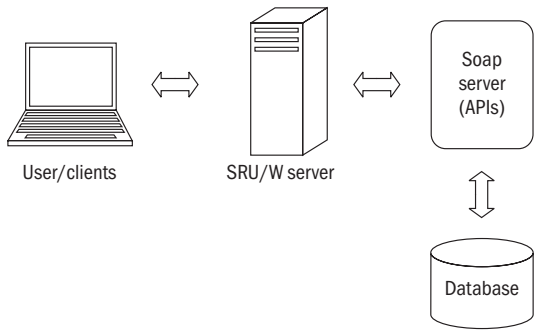


Figure1 Interaction of different modules

Further information is provided as follows.

SRU/W server

The SRU/W server facilitates M2M interaction between clients and the server. Index Data's SimpleServer (http://www.indexdata.dk/simpleserver/) is used to build the SRU/W server. This is a Perl module intended to develop Z39.50, SRU and SRW servers over any type of database. It is based on the popular YAZ toolkit and is robust, efficient, portable, and interoperable with all Z39.50, SRU and SRW clients. The server involves a SOAP envelop to wrap and transfer data. This implementation allowed us to have a distributed environment.

SOAP server

HILT uses a SOAP server to interact with the database and wrap the results in SKOS. REST is a recent alternative to SOAP; the project

[36] http://www.jisc.ac.uk/aboutus/strategy/strategy0709/strategy_aims.aspx

evaluated both options and decided on SOAP because it is more stable. This implementation is based on simple and easy-to-use NuSOAP (a group of PHP classes), which helps to create and consume SOAP web services based on SOAP 1.1, WSDL 1.1 and HTTP 1.0/1.1. It does not require any special PHP extensions. By using SOAP, it was possible to encapsulate database access through a few functions. It is also possible to replace SOAP with REST at a future date if the need arises. In this case, HILT functions would require to be implemented in REST.

Database

HILT data is stored using SQL Server 2005 database software and contain 12 Vocabularies[37] (AAT, CAB, GCMD, HASSET, IPSV, JACS, JITA, LCSH, MeSH, NMR, SCAS and UNESCO) separate from DDC. Tables are designed in such a way so as to accommodate future versions of vocabularies, and flat tables are also designed to improve performance. Interaction with the database is through stored procedures, which adds a layer of security and better performance. A full-text index facilitates the search and retrieval of data efficiently and SQL Server's ability to rank result records (based on the frequency of search terms) is useful in multi-term queries. Database access is supported through the SOAP server.

APIs

As part of the HILT toolkit we have various APIs to offer, which are accessible using the SRW/U server. Results are wrapped in SKOS (Simple Knowledge Organisation System).

The APIs are as follows.

1 *Get_DDC_records* Returns DDC captions and numbers related to a subject term. The user can then choose the most appropriate to his/her interest. This function searches within DDC and any schemes that are mapped to DDC for an input term and returns only DDC instances.

2 *Get_collections* Returns collections classified under a specified DDC number or its stem, including subject scheme used. A process of truncation is used within this function in order to maximise retrieval potential and avoid a 'no hits' situation. For example, if a user searches for collections with DDC number 336.26, HILT returns collections matching 336.26, 336.2, 336, 330, and 300.

3 *Get_non_DDC_records* Returns terms from schemes other than DDC by matching user terms to DDC notations, before identifying mappings to those particular numbers. This function also employs truncation to retrieve relevant mappigs. For an input 336.36, HILT returns mappings corresponding to 336.36, 336.3, 336, 330 and 300. It is up to the client to choose the mappings of their interest.

4 *Get_all_records* This function combines the output of get_DDC_records and get_non_DDC_records.

5 *Get_filtered_set* Get_filtered_set enables a user to search a particular scheme or schemes directly for a specific term, together with its broader, narrower, related and non-preferred terms, if selected and where applicable. You can also search for the id of a term in a scheme (if known), to build a hierarchy of that scheme where available.

6 *Get_SP_suggestions* HILT Spell Checker can suggest a list of words similar to the input word. This implementation is based on David Spencer's code using the n-gram method and the Levenshtein distance. An index (the dictionary) with all the possible words from the HILT database (a lucene index) is created, and suggestions are based on this index. The index can be extended by including terms from other sources as well.

[37] Plus the DDC spine

A single term query returns multiple records, whereas a multi-term query returns a single record.

7 *Get_wordnet_suggestion* Returns the description about the input term from a large lexical database of English - WordNet. Implementation is based on JwordNet - Java interface to WordNet.

Toolkit

A toolkit has been designed to illustrate how these various APIs can be embedded within a service and provide intending implementers with a basic 'starter' package. Further details are available at: http://hilt4.cdlr.strath.ac.uk/toolkit/intro.cgi. A Perl version is available to download and a PHP version may be available at a later date.

Possible uses

As indicated earlier, a description of how the functions and the toolkit may be used in practice is provided in McCulloch (2008). At present, these are pilot facilities, but the aim is to move gradually towards operational services, as described in the appropriate section below. Operational facilities will enable national, institutional, and other information service providers to access and use terminological and interoperability data to enhance their own services in a variety of ways, including, but not necessarily limited to, the following.

1 Improving recall in a subject search of one or more databases by enriching the set of terms known to a user by providing synonyms and related terms.
2 Providing the best terms for a subject search in a remote service that uses a subject scheme unfamiliar to 'home service' users (or in a cross-search of a group of such services).
3 Taking a user's subject term and using it to identify relevant information services via registries such as IESR (http://iesr.ac.uk/).
4 Generating an interactive browse structure where a scheme is arranged hierarchically.
5 Offering the ability to send a term from a chosen subject scheme and receive back data on broader terms, narrower terms, hierarchy information, preferred and non-preferred terms, and so on.
6 Providing cataloguing staff with information on subject schemes and inter-scheme mappings to assist in metadata creation.
7 Providing a spell-check mechanism to assist user searching.
8 Providing a service to assist user search formulation by providing information on search terms entered (for example, what the term means, alternative meaning, synonyms of the word so that it might be useful in a search, and so on).

Uses of this kind can, of course, already be tested using the pilot facilities currently in place. Some work in this area has already taken place through two small elements of embedding work undertaken within the core HILT IV project; further work is already underway within the embedding extension project.

The HILT architecture and its implications

A key element of the current state of play with the project is developing understanding of the architecture required to support a rich and user-adaptive set of relevant terminology and subject interoperability services, and the implications of this architecture for a distributed and collaborative approach to future service growth and associated research and development.

Beginning with JISC-funded work with UK user communities in the first phase of HILT, the project's original focus as regards solving subject interoperability problems was on mapping between subject schemes via a DDC

spine. Mapping is an established and effective approach (for example, Mayr and Petras 2008) and has attracted significant resourcing in other European countries (for example, Agosti *et al.* 2007; Mayr and Petras 2008). However, HILT now has an architectural approach that allows it (1) to adopt a range of interoperability strategies appropriate to specific use cases and cost-benefit levels (for example, expensive deep mapping where the problem and user area justifies this, or high-level or browse-based retrieval where significant costs would be less justifiable), (2) to incorporate in the model - for the benefit of the JISC user communities - solutions offered and funded by other players, whether they be based on mapping to a spine, scheme to scheme, or some variety of automated approach.

The original assumption underpinning HILT was that it would be, in essence, a stand-alone service that would facilitate subject interoperability in the JISC Information Environment by mapping commonly used schemes to a DDC spine and providing the best terms to use in services using a particular subject scheme via these mappings. All through the progress of the project, it was very clear (1) that, even within the JISC communities, this was likely to be a large and probably expensive task (2) that mapping to a DDC spine and, indeed; mapping in itself was only one of the many possible approaches to the problem, and (3) that the variety of subject interoperability problems within JISC would probably be better tackled using a mix of these methods. However, the likely additional costs entailed in encompassing a variety of approaches in a HILT service meant that only one approach – the aforesaid mapping via DDC spine – could be encompassed in a project with limited resources.

Fortunately, the move towards HILT offering M2M functionality has changed the picture somewhat. In the kind of web service environment now envisaged, it becomes possible to think in terms of a distributed and devolved approach to subject interoperability described in Appendix A (Nicholson 2008, for a more detailed description)—with information services, both within JISC and elsewhere, utilising M2M connections, not just to HILT, but to terminology services worldwide, as they come on stream. These would be funded by a range of different players, and would utilize all kinds of different approaches to interoperability. Information services would find and interact with them via data obtained through infrastructural services such as IESR or dedicated terminology services registries. Other data from the same sources would allow the information services to elegantly handle the different kinds of data and different approaches to interoperability served up by these different terminology services.

With this as the backdrop, it becomes necessary to look at the question of research required for subject interoperability service development in the context of the distributed and devolved architecture it implies. The downside of this is that a research for development agenda that is more complex and wider in scope than might otherwise be the case (since it must handle not just interactions with one service of known functionality and scope, but (1) a large variety of different services of unknown functionality and scope, and (2) intermediate interactions with infrastructural services such as registries). The upside is that the distributed and devolved environment means (1) that JISC need only focus on those aspects of research for development needed by specific JISC communities at specific times (2) that other players in the field will likely seek to tackle elements of the research for development agenda required in their own domains, and (3) it is probable that some of the issues faced

in the JISC communities will also be faced (and, in some cases, tackled) elsewhere.

This still emerging perspective on the likely technical and networked environment within which a future service will grow will impact on the upcoming considerations relating to the shape and form of a sustainable service and on probable R&D paths pursued.

Project understanding of service and end-user needs

At present, HILT has only a limited understanding of the needs of the services likely to use HILT facilities and an even more limited understanding of the needs of the end users of such information services. Work on embedding HILT facilities transparently into the user interfaces (or clones of those interfaces) of three services has been undertaken: the GoGeo! Service, which used HILT terminological data to enrich user search terms sets when searching distributed external services external to GoGeo!; the Intute service, which used HILT facilities in spell-check and in offering users a drop down list of terms related to the term input (for example motherhood and mother and child relationships as alternatives to mothers); and the CAIRNS distributed catalogue service, which established a proof-of-concept demonstrator to show how HILT could be used to improve recall. These were small-scale exercises, but nevertheless provided useful insights into the kinds of issues likely to be encountered as HILT aimed to offer its services operational to a range of disparate services with different aims and different terminological issues (it is safe to say that almost every information service operates in a unique subject terminologies environment making 'plug and play' solutions to all but a few core problems difficult to provide). They also helped inform the need for the embedding extension project and, to some extent, its shape and form.

HILT IV also carried out some limited work with a number of individuals who joined a HILT email list (HILT-Collaborators@jiscmail.ac.uk) in response to a call for service staff who were potentially interested in using HILT pilot facilities. This work showed that there was a good deal of interest from the community in the kind of uses of HILT listed above. Twenty-nine people joined the list which was set up specifically to look at the embedding work of this kind. A questionnaire asking what kinds of uses were of interest was answered by 16 people. All indicated an interest in at least one of the above potential ways of using HILT, most indicated an interest in three or more ways. However, the work also showed that in many cases – via some practical tests of actual training and attempts to do simple embedding work – it would be difficult for the early stages of embedding work using HILT facilities to be supported remotely and that more work was needed to discover the best way of ensuring that embedding work at service sites was facilitated and well supported by HILT.

This work notwithstanding, however, it is safe to say that further in-depth work with both staff of services interested in using HILT's M2M services to support subject searching and interoperability functions within their service interfaces, and, more particularly, with the end-users of such information services is needed to inform the process of transforming pilot facilities into a robust, useful, and sustainable service. Some of this work is scheduled to take place within the embedding extension to HILT IV, but a good deal of it will have to be encompassed within a later scoping study for a future service, and in a projected subsequent 'soft launch' of an actual service.

The requirements of a robust, useful, and sustainable service

As previously noted, one of the outcomes of HILT Phase IV was the conclusion that the

best general option for a sustainable operational Shared Infrastructure Service based on HILT project outcomes was a Terminologies Interoperability Centre. More precisely, it was determined - based on an understanding of issues in the area of subject interoperability as developed over a number of phases of HILT, and a discussion with JISC itself on how to best fit future developments into JISC's strategic requirements and also provide a useful, reliable, and sustainable service for JISC-related communities - that the best route forward was to work towards the 'soft launch' of a Terminologies Interoperability Centre. Once launched, this would offer the community a mix of free and charged-for services and would be supported by a mix of JISC funding, externally earned income and R&D funding, and collaboration, and would provide the following,

- M2M and user-level access to terminology sets, the detail of those terminology sets, and data to facilitate interoperability between them.
- Open source software toolkits that would enable M2M interaction with HILT web services to be transparently embedded in the user interfaces of local, national, and project information services.
- A basic architecture for terminology and interoperability services in the JISC Information Environment (and potentially beyond).
- A way of mounting and developing new terminologies and terminologies interoperability data required by the community, including JISC-specified work to facilitate improvements in subject access in and between the various JISC user communities and their external partners based on ongoing assessments of user and service needs.
- Advisory and M2M support services for projects, services, and other initiatives in JISC or JISC institutions where there is a subject description, subject retrieval, or subject interoperability facet.
- A JISC funded free advisory and training service on using the above facilities in local or national services and projects in the percentage of cases where this was relatively straightforward (plug 'n' play, but after a bit of advice and training).
- The development and hosting of a JISC-focused terminologies services registry.
- A charged-for consultancy service where the work and advice required by local and national services, projects, and organisations (both within and out of JISC) was less straightforward or more sophisticated (because of unique client and client service circumstances and terminology sets).
- A portal for tools and training in the areas described above.
- A focus for wider work in the terminologies area, funded through a variety of sources, including non-JISC sources (for example though successful bids for European funding).
- Ongoing work to facilitate JISC involvement and leadership in a strategically important area with significance for both subject-based retrieval needs in research, learning, teaching, and and semantic web developments.

The initial plan in respect of this was to move directly from HILT Phase IV to the proposed soft launch. However, initial consideration of this proposal quickly proved that an interim stage - a scoping study for the proposed Centre and its services – was a necessary and sensible first step. If funded, this scoping study will provide a well-researched evidence base that will inform and guide a future 'soft launch' of a Terminologies Interoperability Centre by

- Putting in place service quality infrastructure to support the work of the

Centre, including further development and testing of the components from HILT IV and work on a pilot terminology services registry. This will ensure that the standard services offered at the soft launch will be robust and usable in a range of JISC service and user environments.
- Determining service user and end user needs via iterative feedback from hands-on experience, utilizing outcomes in TIC scoping and soft launch plans, creating mechanisms for an ongoing assessment of such needs, and identifying specific players to work with TIC during the soft launch period.
- Scoping in detail what free and charged-for services the Centre should offer and what they would cost.
- Producing a bid for TIC start-up costs, a programme of works, and a well-researched Sustainability Plan.

The proposal is that this study would take place between June 2009 and November 2010, with an actual soft launch of the Centre taking place in December 2010.

The embedding extension project and future research and development directions

This, then, is the position with the project at the end of the core element of Phase IV. The next steps for HILT are, first of all, to conduct the more complex embedding work planned under the embedding extension project, then - assuming a positive outcome to the proposal for a Terminologies Interoperability Centre scoping study to begin to determine the best shape and form for a future scalable and sustainable service. Following that - and regardless of the outcome of the scoping study proposal the logical path to follow is one that will lead towards increasing international collaborative work based on an agreed architecture and common R&D programme.

Future: the embedding extension project

As indicated earlier, a small extension project was funded towards the end of HILT Phase IV with the aim of embedding the pilot use of HILT M2M terminology and interoperability facilities within the user interfaces of a number of information services in order to enhance functionality for the end-users of these services. The services in question are based at EDINA, Mimas (Intute), and CDLR. A second extension project, funded in parallel at IESR in Manchester, also entailed HILT services embedding plans and collaboration between the two with respect to HILT embedding is taking place. The project runs from January to May 2009. It is underway at the time of writing, but has yet to produce significant results with respect to embedding as such (although some associated mapping work between DDC and JACS and DDC and RAE headings has been completed).

The focus of the combined CDLR and IESR work as presently envisaged is as follows.

IESR

The aim will be to use HILT M2M services to enhance the user browse interface to the IESR collection discovery service in a variety of ways by sending user terms to HILT, and enrich the interface by getting back broader, narrower, and related terms associated with various HILT schemes and using them to improve retrieval.

Intute

Intute is currently being redesigned, with a target date of end of July. This means embedding demos would not be sustained much past the end of April. Accordingly, it was suggested that Intute work on a clone of their service to show what HILT can offer—the likely focus being enhancing user searches in various ways (for example, by offering more specific terms when a user puts in a very broad one (like 'farming').

Edina

Edina will aim to improve their Depot[38] service by using HILT to guide user-provided subject tags via an M2M fed drop down list of standard terms (for example, from JACS). They will also look to improve their VSM service by aiming to provide users with term expansion—taking a user input search term and expanding it via HILT to include terms from multiple schemes.

CDLR

The intention here is to enhance the recently developed Google Maps based Scotland's Information[39] service by using HILT to translate existing subject terms into the same subjects in different languages. Languages being considered are those where HILT has some access to DDC captions in the language concerned and can therefore cross-map - German, French, and Welsh-Gaelic.

In addition to the above, there will also be some generic work on optimizing the HILT spell check pilot mechanism for use in UK HE an FE.

Future: the TIC scoping study

The aim of the Scoping Study is to provide a well-researched evidence base that will inform and guide a future 'soft launch' of a Terminologies Interoperability Centre by scoping out the detail of a sustainable mix of JISC-funded and charged for services designed to help meet JISC strategic aims with respect to serving its stakeholders. The basic outline of what is required with respect to addressing subject interoperability issues is known. What is needed now is a clear, evidence-based indication of what is required in detail to ensure a robust, sustainable service sensitive to community needs and the strategic aims of the JISC. Amongst other things, this requires intensive work in which specific JISC-related user communities (including both service and end users) are given hands on experience of the terminological tools as they become available as part of an iterative and ongoing approach to understanding, and implementing solutions to, user needs in specific communities.

The intention is to work with following four target groups.

1 HE and FE libraries (subject librarians, cataloguers/repository managers, systems librarians/repository developers, end-users of various kinds, national service providers.
2 E-learning (teachers, learners, librarians, VLE managers, VLE developers, national service providers.
3 NHS e-library (and their services).
4 A cross-domain group entailing Museums, Archives, Libraries, and so on, not just in HE (for example, public libraries as well as HE libraries and work with the Strategic Content Alliance)

Working with these groups, the project will move gradually and interactively towards a reliable understanding of user requirements via the following threads:

- Four one-day workshops at four UK sites with hands-on work, focus groups, and other activities designed to engage and inform, and set up an ongoing iterative process to allow improved understanding of needs, with each to have a mix of the target groups, and be shared with IESR.
- Workshop demonstrations of terminology service embedding possibilities and uses and discussions on approaches, problems, and issues aimed at finding out a more about real needs.
- A purpose built social networking site would be set up and populated prior to the workshops and tools showcase added to it, creating a 'virtual lab' for ongoing interaction with user groups.

[38] http://depot.edina.ac.uk/
[39] http://www.scotlandsinformation.com/

- During and after the workshops, the environment would be used to show possibilities and get feedback aimed at pinpointing work needed in a soft launch; we would also use it during the soft launch.
- Discussions with these groups would be mainly on their own experiences with subject searching and terminology issues but would also go beyond terminology issues to discuss issues like the availability of local expertise, and how we might best pitch (1) an 'off-the-shelf' (and 'free') service (2) how we might cost a paid for consultancy (something that would also be discussed with consultants in the field). There would also be discussions on protocol, mark-up languages (SKOS, MARC, Zthes), and specific mapping and tools functionality needs.
- Work with the groups during and after the workshops to show them illustrative functionality and terminological improvements based on ongoing interaction.
- Work on the detail of how the chosen groups (or individual services or organisations/institutions from these groups) might be involved in a direct practical sense in a soft launch and what level of funding would be required by them for their involvement.

Future: towards a common collaborative research and development programme

As indicated above under The HILT Architecture and its Implications, HILT sees a need, both within the JISC community, and in the world at large, for a globally-scoped programme of collaborative R&D based on a common view of an inclusive architecture for subject interoperability service design. Although HILT progress in this area is currently limited, efforts have nevertheless been made to begin work towards agreeing a collaborative approach with other 'players' in the terminologies field. A paper on the idea was presented at the Ontologies Conference in Helsinki in November 2007. A paper on the topic published in the international version of the *Signum* journal in November 2008, and, more recently (December 2008), steps have been taken to contact major European projects in the terminologies area to begin the process of talking about collaboration and about applying for FP7 funding to carry the work forward.

Discussions are at a very early stage, but if a proposal were put forward, it would be based on the idea that despite the plethora of different means of tackling inter-KOS and inter-lingual interoperability currently in play and the size and complexity of the problem itself, it is both desirable and possible to facilitate gradual cumulative progress towards optimizing subject interoperability across the networked world through geographically distributed co-ordinated collaborative effort – that it can be done by agreeing a model set of requirements for interoperability service design and collectively pursuing a common R&D agenda based on it.

A key aim would be to bring together major (but currently stand-alone) European projects working in the area, including HILT, MACS, KoMoHe, CACAO, and others. If funded, the project would aim to kick-start and give strategic direction to the major and long term programme of work required to research and develop (1) the creation of a pan-European (and ultimately global) web-services-based M2M architecture to support the level of inter-KOS and inter-lingual subject interoperability required to support such things as seamless access to distributed European digital resources for all EU citizens and a richer inter-initiative semantic web; (2) the wide and complex range of user interface handling routines that individual European information services would require to embed

in their interfaces to enable them utilize distributed terminology and terminology cross-walk services transparently for the benefit of their users seeking to search external information services; (3) The ontology of the key elements of the domain and their inter-relationships that would be required to drive these user interface routines, uniquely categorizing such things as KOS types, KOS cross-walk types (for example, based on intellectual mapping via different spines, direct scheme to scheme mappings, automated statistical mappings, clustering, autoclassification), services characteristics, user characteristics and other elements implied by the architecture shown below; (4) an ongoing infrastructure to support future R&D efforts over the long term; and (5) mechanisms to support person-specific, user-group-specific, and task specific customization of system responses to service requests.

It is by no means certain that such a proposal will materialise in the event. However, a number of key players have at least indicated a clear interest in the idea.

Conclusion

The Phase IV project has taken HILT to the point where the launch of an operational support service in the area of subject interoperability is a feasible option and where both investigation of specific needs in this area and practical collaborative work are sensible and feasible next steps. Moving forward requires detailed work, not only on terminology interoperability and associated service delivery issues, but also on service and end user needs and engagement, service sustainablity issues, and the practicalities of interworking with other terminology services and projects in the UK, Europe, and global contexts.

Acknowledgement

Acknowledgements are due to George Macgregor who contributed to HILT IV during the early part of its duration, to James Farnhill and Tom Franklin for providing helpful JISC support, Kathleen Menzies for help with references, OCLC for ongoing support and advice, and to JISC for funding HILT.

References

Agosti, et al. 2007
Roadmap for MultiLingual Information Access in the European Library
In *Research and Advanced Technology for Digital Libraries,* Vol. 4675/2007
Berlin/Heidelberg: Springer
Abstract online at <http://www.springerlink.com/content/g8126155w7518333/>

Apps A. 2007
Using an Application Profile Based Service Registry
In *Proceedings of the International Conference on Dublin Core and Metadata Applications* (*DC 2007: Application Profiles: Theory and Practice*), edited by S A Sutton, S C Chaudry, C Khoo
Singapore, 27–31 August 2007, pp 63–73. Details available at <http://epub.mimas.ac.uk/papers/2007/dc2007/apps-dc2007.html>

Ardo A. 2004
Automatic Subject Classification and Topic Specific Search Engines – Research at KnowLib
Presented at DELOS Regional Awareness Event: Between Knowledge Organization and Semantic Web: Semantic Approaches in Digital Libraries
Lund, Sweden. Details available online at <http://www.delos.info/eventlist/LUB1/Anders_Ardo/DELOS.PDF>

Binding C, Tudhope D, and May K. 2008
Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction via the CIDOC CRM
Proceedings (ECDL 2008) 12th European Conference on Research and Advanced Technology for Digital Libraries Aarhus, 280–290
Lecture Notes in Computer Science, 5173
Berlin: Springer. Final preprint presentation

Blair D C. 1980
Searching biases in large interactive document retrieval systems
Journal of the American Society for Information Science, **31**(4), 217–277

Brown M E. 1995
By any other name: accounting for failure in the naming of subject categories
Library and Information Science Research. **17**(4), pp. 325–429

Buckland M K, Chen A., Gebbie M, Kim Y and Norgard B. 2000
Variation by Subdomain in Indexes to Knowledge Organization Systems
Presented at International Society for Knowledge Organization, ISKO6, Toronto, 2000
Details available at <http://metadata.sims.berkeley.edu/papers/iskopaper.html>

Cacao. 2008
Cross-language Access to Catalogues and Online Libraries partners
CACAO Presentation. Details available at <http://www.cacaoproject.eu/fileadmin/media/presentations/CACAO_presentation.pdf>

Chan L and Zeng M. 2002
Ensuring interoperability among subject vocabularies and knowledge organisation schemes: a methodological analysis. 68th IFLA Council and General Conference, 18–24 August 2002
Glasgow: Details available at <http://www.ifla.org/IV/ifla68/papers/008-122e.pdf>

Clavel-Merrin G. 2004
MACS (Multilingual Access to Subjects): A Virtual Authority File Across Languages
In *Cataloguing and Classification Quarterly*, **39**(1/2)
Berlin: Springer

Day M, Koch T, and Neuroth H. 2004
Searching and browsing multiple subject gateways in the Renardus Service
In *Recent developments and applications in social science methodology: proceedings of the Sixth International Conference on Logic and Methodology* edited by C van Dijkum, J Blasius, Kleijer H, and B Hilten van B. 17–20 August 2004
Amsterdam*:* Amsterdam: SISWO Instituut voor Maatschappijwetenschappen. (CD-ROM). Details available at <http://www.ukoln.ac.uk/metadata/publications/rc33-2004/renardus-paper.pdf>

Doerr M. 2001
Semantic problems of thesaurus mapping
Journal of Digital information, **1**(8). Details available at <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>

Geser G. 2008
STERNA Technology Watch Report. Full Report
Ref: STERNA Del.6.5**,** 10 December 2008. Salzburg Research
Details available at <http://www.sterna-net.eu/index.php/en/downloads>

Godby C J, Miller E J and Reighart R R. 1999
Automatically Generated Topic Maps of World Wide Web Resources
Annual Review of OCLC Research, 1999. Details available at <http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000002655:000000059193&reqid=9300>

Golub K and Tudhope D. 2008
Terminology Registry Scoping Study (TRSS): Excerpt on Metadata
Details available at <http://www.ukoln.ac.uk/projects/trss/dissemination/metadata.pdf>

Hillman D, Stuart A S, Phipps J, Ryan L. 2006
A Metadata Registry from Vocabularies UP: The NSDL Registry Project, submitted to Dublin Core 2006 Conference.
Details available at <http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0605111>

Koch T and Vizine-Goetz D. 1998
Automatic Classification and Content Navigation Support for Web Services: DESIRE II Cooperates with OCLC
Annual Review of OCLC Research, 1998. Details available a <http://www.lub.lu.se/tk/demos/class-ws/automatic.htm>

Landry P. 2004
Multilingual Subject Access: The Linking Approach of MACS
In *Cataloging & Classification Quarterly*, **37**(3/4)

Levergood B, *et al.* 2008
The Specification of the Language of the Field and Interoperability: cross-language access to catalogues and online libraries (CACAO)
International Conference on Dublin Core and Metadata Applications
Berlin. Details available at <http://dcpapers.dublincore.org/ojs/pubs/article/viewFile/933/929>

Liang A C and Sini M. 2006
Mapping AGROVOC and the Chinese agricultural thesaurus: definitions, tools, procedures
New Review of Hypermedia and Multimedia **12**(1), pp. 51–62

Macgregor G and McCulloch E. 2006
Collaborative tagging as a knowledge organisation and resource discovery tool
Library Review **55**(5): pp. 291–300
Pre-print available at <http://strathprints.strath.ac.uk/2335/1/strathprints002335.pdf>

Mayr P and Petras V. 2007
Building a terminology network for search: the KoMoHe project
In Proceedings of the International Conference on Dublin Core and Metadata Applications 2008.
Details available at <http://arxiv.org/pdf/0808.0518>

Mayr P and Petras V. 2008
Cross-concordances: terminology mapping and its effectiveness for information retrieval
IFLA 2008 Conference Proceedings. Details available at <http://eprints.rclis.org/13828/>

McCulloch E, Shiri A, and Nicholson D. 2004
Subject searching requirements: the HILT II experience
Library Review **53**(8). pp. 408–414

McCulloch E. 2008
Developing a pilot toolkit to improve subject interoperability between collections and services within the JISC Information Environment
JISC Information Report

McCulloch E and Macgregor G. 2008
Analysis of equivalence mapping for terminology services
Journal of Information Science, **34**(1) pp. 70–92

McCray A T, Loane R F, Browne A C, Bangalore A K. 1999
Terminology Issues in User Access to Web-based Medical Information
Details available at <http://www.amia.org/pubs/symposia/D005626.PDF>

Nicholson D and McCulloch E. 2006a
Investigating the feasibility of a distributed, mapping-based, approach to solving subject interoperability problems in a multi-scheme, cross-service, retrieval environment
International Conference on Digital Libraries, 5–8 December 2006
New Delhi: Details available at <http://strathprints.strath.ac.uk/2875/1/strathprints002875.pdf>

Nicholson D and McCulloch E. 2006b
HILT Phase III: Design requirements of an SRW-compliant Terminologies Mapping Pilot, 5th European Networked Knowledge Organization Systems (NKOS) Workshop, 10th ECDL Conference, 21 September 2006
Alicante, Spain

Nicholson D. 2008
A Common Research and Development Agenda for Subject Interoperability Services?
Signum, **5**

Shiri A, Nicholson D, and McCulloch E. 2004
User Evaluation of a Pilot Terminologies Server for a Distributed Multi-scheme Environment
Online Information Review, **28**(4), 273–283

Shiri A. 2008
Metadata-enhanced Visual Interfaces to Digital Libraries
Journal of Information Science **34**(6), 763–775

Stafford A, Shiri A, Ruecker S, Bouchard M, Mehta P, Anvik K, Rossello X. 2008
Making Sense of Interface Design: implementing searchling, a visual thesaurus-enhanced interface for multilingual digital libraries
Chicago Colloquium on Digital Humanities and Computer Science, 1–3 November 2008
Chicago: University of Chicago

Teevan J, Alvarado C, Ackerman M S, Karger D R. 2004
The Perfect Search Engine is not Enough: A Study of Orienteering Behaviour in Directed Search
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '04)

Tudhope D. 2006
A tentative typology of KOS: towards a KOS of KOS?
5th European NKOS Workshop, European Conference on Digital Libraries (ECDL06)
Alicante. Details available at: <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2006/papers/paper-018.doc>

Tudhope D, Koch T, and Heery R. 2006
Terminology Services and Technology JISC state of the art review
Details available at: <http://www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf>

van Gendt M, *et al.* 2006
Semantic Web Techniques for Multiple Views on Heterogeneous Collections: a case study
Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006), edited by Julio Gonzalo, Constantino Thanos, M. Felisa Verdejo and Rafael C Carrasco
LNCS, **4172,** pp. 426–437, Alicante: Springer Verlag. Details available at: http://www.cs.vu.nl/~mtvgendt/STITCH-ECDL06.pdf

Vizine-Goetz, Diane, Andrew Houghton, and Eric Childress. 2006
Web services for controlled vocabularies
Bulletin of the American Society for Information Science and Technology, **35**(5)
Details available at <http://www.asis.org/Bulletin/Jun-06/vizine-goetz_houghton_childress.html>

Zeng Marcia Lei and Lois Mai Chan. 2004
Trends and issues in establishing interoperability among knowledge organization systems
Journal of American Society for Information Science and Technology **55**(5): pp. 377–395

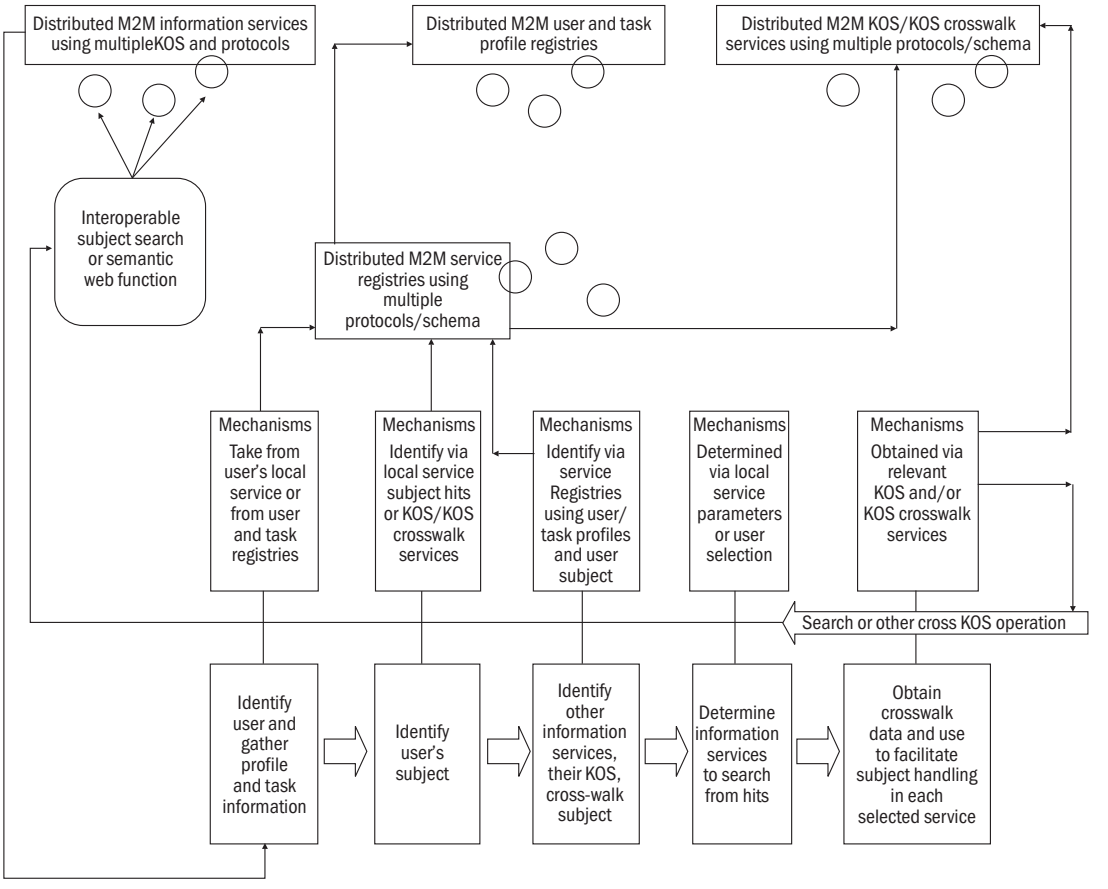
Zeng M L and Chan L M. 2006.
Metadata interoperability and standardization – a study of methodology Part II: Achieving interoperability at the record and repository levels
In *D-Lib Magazine* **12**(6)
Details available at <http://www.dlib.org/dlib/june06/zeng/06zeng.html>

Appendix A: the HILT architecture

The diagram below shows the HILT architecture in outline.

In a working model, the following are true.

- There are multiple service registries available that 'know' about each other, so any given information service only needs to know about their own 'home' service registry
- There are multiple information services available, recorded in and discoverable via one or more service registries, and classified according to their subject coverage
- There are multiple user and task profile registries available, also recorded in and discoverable via one or more service registries
- There are multiple KOS and KOS 'cross-walk' services available, also recorded in and discoverable via one or more service registries (or possibly terminology registries)
- There are many information services available using many different KOS and access protocols
- The user's start point is usually (but not necessarily) one of these information services
- The users home information service offers the user a facility to cross search other information services appropriate to her subject needs and uses a subject interoperability service (encompassing the whole of the diagram above) to underpin this service.
- Most of the workings of this service are transparent to the user
- The assumption is that the user starts at one of the information services and needs to search one or more (possibly unknown) others
- The home service goes through the processess shown in the five boxes along the bottom of the diagram, using the services registries, user and task profile registries and KOS and KOS crosswalk services shown in the upper half of the diagram as necessary



- In order to elegantly and intelligently (and transparently) handle often complex interactions as described above, a local information service would need access to complex metadata and sophisticated embedded user interface routines to handle the interactions

Appendix B: Glossary

AAT	Art and Architecture Thesaurus
API	Application Programming Interface
CAB	Commonwealth Agricultural Bureaux
CAIRNS	Co-operative Information Retrieval Network for Scotland
CDLR	Centre for Digital Library Research
DDC	Dewey Decimal Classification
Depot	A UK national open access repository for researchers not yet having an institutional repository in which to deposit their papers, articles, and book chapters (e-prints)

EDINA	A JISC-funded national datacentre based at Edinburgh University Library, offering the UK tertiary education and research community networked access to a library of data, information and research resources.
FE	Further Education
GCMD	Global Change Master Directory
Go Geo!	A tool designed to help users find details about geo-spatial datasets and related resources within the UK tertiary education sector and beyond. A trial service is provided by EDINA.
HASSET	Humanities and Social Science Electronic Thesaurus
HE	Higher Education
HILT Project	High-level Thesaurus Project
IESR	JISC Information Environment Service Registry
intute	intute is a free online service providing access to the very best web resources for education and research. Formerly the Resource Discovery Network (RDN).
IPSV	Integrated Public Sector Vocabulary
JACS	Joint Academic Coding System
JISC	Joint Information Systems Committee
JISC IE	Joint Information Systems Committee Information Environment
LCSH	Library of Congress Subject Headings
M2M	Machine to machine interaction
MARC	Machine readable cataloguing
MeSH	Medical Subject Headings
NHS	National Health Service
NMR	National Monuments Records Thesauri
OCLC	Online Computer Library Center
RAE	Research Assessment Exercise
REST	Representational State Transfer
RIN	Research Information Network
SCAS	Standard Classification of Academic Subjects
SKOS	Simple Knowledge Organization System. SKOS Core supports the Resource Description Framework (RDF) description of language-oriented knowledge organisation systems (KOS), such as thesauri, glossaries, controlled vocabularies, taxonomies and classification schemes
SOAP	Originally the Simple Object Access Protocol, but now more simply referred to as SOAP. Used to exchange XML-based messages over computer networks, normally using HTTP
SQL	Structured Query Language
SRU/W	Search/Retrieve Web Service and Search & Retrieve URL - Z39.50 Next Generation
UKOLN	A centre of expertise in digital information management, providing advice and services to the library, information, education and cultural heritage communities. Based at the University of Bath and formerly known as the UK Office for Library & Information Networking
UNESCO Thesaurus	United Nations Educational, Scientific and Cultural Organization subject scheme

Use Case	A Use Case represents a series of interactions between a user (human or machine) and the system, utilising (in the present case) an M2M link. Typically, the interaction starts with an enquiry and leads to a resource that should answer that enquiry
VLE	Virtual Learning Environment
WSDL	Web Services Description Language
XML	Extensible Mark-up Language
Z39.50	An international standard specifying a client/server-based protocol for searching and retrieving information from remote databases
Zthes	The Zthes profile is an abstract model for representing and searching thesauri and specifies how this model may be implemented using the Z39.50 and SRW protocols

Digital library infrastructure in Finland – political decisions empowering development

Kristiina Hormia-Poutanen
Deputy National Librarian, Director of National Library Network Services
National Library of Finland, P.O.Boc 26 (Teollisuuskatu 23–25)
00014 University of Helsinki

World Digital Libraries 2(1): 23–32

Abstract

This paper presents an overview of effects of recent political decisions on the higher education network as well as the research environment in Finland and in particular describes their impact on digital library services. The driving force behind the government decisions and actions aim to strengthen internationally competitive research and innovation in the country. The change of the legal status of universities and restructuring the higher education network, the realization of Research Infrastructure Survey and Roadmap and the launch of the Finnish Digital Library project are the main actions affecting the development of library services. Particularly in light of the current global economic situation the need for national infrastructures which can be applied at institutional level becomes vital. The cooperation between Finnish libraries and the National Library of Finland is advanced but in the emergence of new digital services there are opportunities for collaborations outside the previous sphere to encompass new partners.

Introduction

The Finnish government defines internationally competitive research and innovation as being the key success factors for society in the long run. Many decisions have been taken, strategies prepared, and projects launched to achieve this aim. This article describes a couple of current projects, which have relevance to the development of library infrastructure.

The government has decided that the higher education network in Finland has to be restructured. A project to develop the network was launched in 2007, and as a consequence, the network is changing. The legal status of universities will change in 2010, and consequently, the universities will become more autonomous than they are now, and economic freedom will increase. At the same time, the higher education network will reduce in size. There have already been some mergers of universities and polytechnics, and more will take place in 2010 and 2011. The aim of these structural changes is to increase the international competitiveness of the Finnish research and improve the quality of research and education. The legal and structural changes also affect library services, and the process of the division of work between libraries and national service providers has to be improved.

The national research infrastructure survey was carried out for the first time in Finland in 2008. As a result, 24 significant national-level infrastructures were identified and 20 proposals were accepted for the roadmap of new needs. Research infrastructure policy will be developed in the future, and it will be linked to national innovation strategy. The emphasis for the development of research infrastructure is on facilitating research of international standards in Finland.

The IT environment of libraries, archives, and museums in Finland is in developing stages. The Finnish Digital Library programme, consisting of two main projects, was launched during 2008 and will run till May 2011. The aim of the initiative is to develop an efficient service for access to digital resources in libraries, archives, and museums, and to develop specifications and recommendations for a national digital long-term preservation system.

The National Library of Finland has been a service centre for libraries since 2006. The library is an active player in all the national projects mentioned above.

Structural changes in the higher education network and their impact on the centralized library IT services

The legal status of universities will change in 2010. The universities will gain more autonomy, and economic freedom and steering by the Ministry of Education will decrease. The major part of the funding for the universities will still come from the Ministry of Education, but the universities will be encouraged to seek additional funding.

For historical reasons, there are as many as 50 institutions of higher education in Finland, a relatively large number for a country with a population of only five million. According to the plan and the recommendations of the Ministry of Education, there will be about 15 universities and 18 polytechnics by 2020. Mergers of higher educational institutions have already started, and more will take place in the coming years. One of the most discussed mergers is the formation of the Aalto University (http://www.aaltoyliopisto.info/en/) in which three universities covering very different disciplines will merge in 2010 (Helsinki University of Technology, Helsinki School of Economics, and the University of Art and Design Helsinki). The driving force in the merger is the idea of encouraging

innovation and new thinking by bringing together education and research in the technical sciences, economics, and art and design.

The Ministry of Education has launched a project to investigate the ways in which the organizational changes taking place at institutions of higher education will affect the libraries. The project will also aim at achieving cost reductions at the library level by strengthening the position of the national library as a central service provider. When planning new services, an appropriate and efficient division of responsibilities among libraries and the National Library will be a top priority.

Digitalization of the working environment of researchers, teachers, and students has been identified as the main issue influencing the development of library services in the next decade. Increasingly, more digital resources will be available through licensing, open access publishing, increasing the availability of e-books, and digitalization on a global level. Research and learning environments are developing, new infrastructure to support learning and research will be created, and easy access to digital content is demanded. It will be very important to be able to provide relevant content for the user at the right phase of the work process and in the right working environment. One of the major drawbacks in the development of the digital working environment is the copyright law. Masses of digitized content are not available to users because of legal restrictions. The needs for research and education should be integrated in the law in the future.

The importance of digital object management systems, digital preservation systems, and systems for easy discovery and delivery of digital content will become fundamental. In a small country like Finland, the only way to produce high quality infrastructure is to do develop it centrally.

Research infrastructure survey and roadmap

In 2008, the Finnish Ministry of Education coordinated a project to identify current national-level research infrastructure and develop a roadmap for new needs. Research infrastructure is the resource for research facilities, equipment, materials and services, permitting research and development at different stages of innovation, supporting organized research, and maintaining and developing research capacity (National Level 2008). Similar work has been carried out in Europe. In 2006, ESFRI, or the European Strategy Forum on Research Infrastructure, published its plan, the so-called roadmap, on the need to construct and update research infrastructure at the European level.

The Steering Group of the Finnish project listed 24 projects out of 184 as significant national-level infrastructures in Finland. One of these is the National Electronic Library (FinELib), which is a joint activity performed by the Finnish libraries and the National Library.

For the roadmap for new research infrastructure, 116 proposals were received from which 20 were selected for the roadmap. System Architecture for Memory Institutions, a proposal representing Finnish libraries, museums, and archives, was one of the accepted proposals.

According to the preliminary estimate provided by the mapping, Finland spends approximately 130 million euros per year in public appropriations for the upkeep of national infrastructure. The construction costs of the projects chosen for the roadmap will be approximately 230 million euros over the period 2008–20, with annual costs for Finland of approximately 30 million euros. Finland needs a centralized funding system for renewing the existing research infrastructure and for funding new projects at the national level.

The centralized funding system should also take into account the need to manage research infrastructure policy and make long-term international commitments.

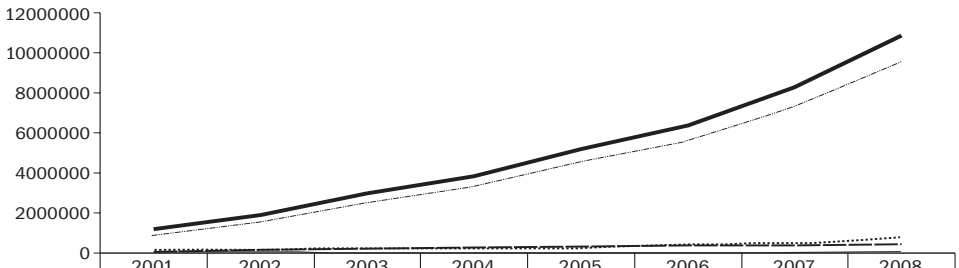
The National Electronic Library, FinELib – a significant national research infrastructure

The National Electronic Library, FinELib, is a national consortium, which supports research and academics in Finland by promoting the availability of high-quality information and its use in society (Figure 1). FinELib acquires most of the electronic resources in Finnish libraries (Hormia-Poutanen 2002b, 2003). The resources available include approximately 20 000 electronic journals and 300 000 e-books plus approximately 130 databases which attract tens of millions of information searches every year (Table 1). The members of the consortium are all Finnish universities, polytechnics, public libraries, and a large number of research institutes.

The acquisitions, which are made through FinELib, cover over 80% of the acquisitions of electronic resources at Finnish universities. Thus, FinELib has a key role in providing electronic materials for the user population of universities. According to user surveys and usage information, the selection of resources meets the needs of the users well. Over the years, FinELib has been operating and showing a growing trend in usage (Hormia-Poutanen 2007). FinELib has been very active in international cooperation related to the development of cost division models, sharing experiences, and giving consultations (Hormia-Poutanen 2006).

The main principles guiding the management of the consortium have been defined in the Memorandum of Understanding. FinELib has a service agreement policy, which covers two main services, namely the licensing of e-resources and the maintenance of the national portal. These service agreements define the responsibilities of the national library and the customer.

The programme is managed through three working groups. The high-level steering group is responsible for policy-making, strategic planning, drawing up the annual plan of action, and evaluating the results. The group consists of top-level management from the universities, polytechnics, research institutes and the assicuated libraries, the public libraries, the Ministry of Education, and the end-users. The consortium group is responsible for more practical issues and consists of library directors from the four library sectors. The expertise of the various fields of science as well as technology issues and the interests of end-users are represented



	2001	2002	2003	2004	2005	2006	2007	2008
Universities	940829	1560359	2542522	3342898	4608491	5597136	7327054	9585419
Polytechnics	149693	162116	192263	204301	263677	406352	508867	745103
Research institutes	36214	140462	202940	296416	322504	357105	372867	450417
Public libraries	42062	28567	16540	13750	14023	22994	45270	69995
Total	1168798	1891504	2954265	3857365	5208695	6383587	8254058	10850934

Figure 1 The trend of downloaded articles in the FinELib consortium

Table 1 Central services for libraries provided by the National Library

Service	*Indicators 2008*
Access to high-quality information National licensing – FinELib	20 000 e-journals available 300 000 e-books 11 million article downloads
National databases	Access to National databases: 10 million searches
Library system development (Voyager)	All universities and polytechnics use Voyager
Portal Development (MetaLib & SFX)	All universities, polytechnics and public libraries use MetaLib and SFX.
Digital Object management system development (Dspace)	All polytechnics and a few universities use Dspace.
Library statistics (KITT)	All universities, polytechnics and several special libraries
National library surveys	All library sectors
Projects Finnish Digital Library: Public interface Structural changes in the libraries of the higher education network Projects related to institutional repositories and open access	

in the expert groups. Their main task is to submit proposals for resources to be licensed in the future and to develop Nelli-portal services at the organizational level. The programme was evaluated in 2003 (Varis and Saari 2003).

Although the final decision has to be taken by the National Library, this three-tier organization guarantees that all consortium members can make their voices heard and influence the decision.

In 2009, FinELib is a well-known and highly valued national research infrastructure. The level of government funding is 4 million euros and the funding is directed towards the acquisition of high-quality electronic resources as well as the development of the national portal. The funding also includes additional costs, such as staff and staff development costs. The estimated total turnover in 2008 was over 16 million euros, consisting of central funding and the organization's own funding.

The Finnish Digital Library – chosen for the research infrastructure roadmap

System architecture for memory organizations was chosen for the roadmap of needed research infrastructures. The Finnish Digital Library forms the main part of the concept. The Finnish Digital Library initiative was launched as part of the Finnish Information Society strategy for 2007–2015. The aim of the initiative is to develop an efficient service for access to digital resources in libraries, archives and museums and to develop specifications and recommendations for a national digital long-term preservation system. The initiative is a very ambitious one. The participating

organizations range from different kinds of libraries to art museums, science museums, audiovisual archives and other archives. Altogether, there are 35 organizations represented in the programme. The IT infrastructure of libraries, museums and archives will be undergoing a major change due to the initiative. The aim is to improve the efficiency of processes, save costs in digitalization, improve access to information and preserve content in the long term for use in education, research and innovation for the benefit of the country.

The Finnish Digital Library project consists of two subprojects, namely the public interface and digital preservation project. The challenges in these two subprojects are different. In developing the digital long-term preservation system, the technical challenges are huge. It is not possible to choose from among a large variety of software. At the moment, the National Library of the Netherlands runs a digital preservation system and the National Library of New Zealand is developing one in cooperation with ExLibris and an international group of experts. In addition, many national libraries have chosen to develop their own digital preservation systems. The experience gained in these projects will be used in the Finnish project.

The challenges of the public interface project are more political or human than technical in nature. Questions such as 'Who is the user' and 'What are the needs of users?', 'What content can be accessed?', 'What is the added value of the new system compared with current systems?' have to be answered.

The needs of users and, particularly, the entrance of the 'Google generation' into universities and other institutions are exerting a powerful influence on the development of library services. The information behaviour of the researcher of the future is described in an interesting and thought-provoking report commissioned by the British Library and JISC (Information Behaviour 2008). Developments in standardisation and IT technology provide libraries and other memory institutions with new tools to react to the new demands. Current library services or the services at museums or archives do not satisfactorily meet end-users' needs because, for example, they are considered too complicated and difficult to integrate into working environments. Additionally, the services do not support the consortia's solutions and divisions of labour.

One of the basic ideas in the development of the public interface is to separate the user interface from the back-end systems. Separating the user interface from back-end systems shifts the developmental focus from libraries to end-user needs. Back-end systems are, for example, library catalogues and other catalogues, institutional repositories, digital preservation systems, learning environments. In the current model, back-end systems also include interface functionality and the needs of end-users and librarians will have to be taken into account (Hormia-Poutanen 2008).

The public interface project will run from 2008 until May 2011 and by then a service for a large group of memory organizations should be up and running. The year 2009 is dedicated to the definition of the functional and technical specifications, preparation of the documentation for a tender and organizing the tender. In 2010, a pilot to test the software(s) will be organized. The National Library of Finland is responsible for the coordination of the public interface project.

The architecture is described in Figure 2.

Background to the current situation

The role of the National Library as a service provider is becoming stronger. New responsibilities, especially related to digital library tasks, are being given to the library. There is a long history behind this

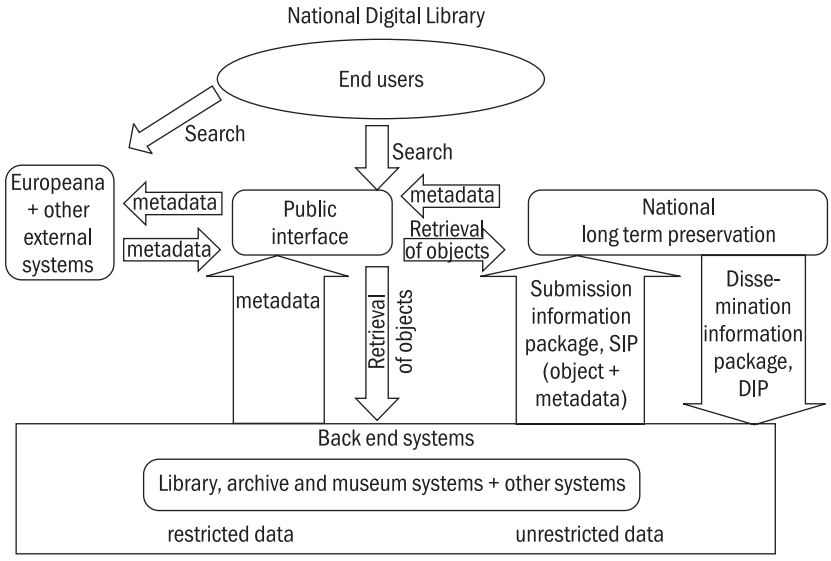


Figure 2 Information architecture for memory organizations

development. The role of the library has grown over the last couple of decades, but the tradition of cooperation among libraries forms an important background to this development.

The Finnish Library Network—cooperation as a working model

The tradition of cooperation among libraries has been an important factor in the development of library services in Finland; the university libraries in particular have a long history of cooperation. One very concrete expression of this cooperation is the shared library system that all Finnish university libraries have been using since the early 1990s. The first integrated library system was called VTLS, but today all university and polytechnic libraries use the Voyager system.

The National Library of Finland's development of national services for libraries has increased the need to improve the coordination and management of each library sector. For the libraries and the National Library it is essential that national service-related negotiations be organized in a timesaving and practical way that avoids repetitive duplication.

University libraries have been organizing cooperation between the libraries in the network for over ten years. The Council of University Libraries <http://www.nationallibrary.fi/libraries/council.html>, formed in 1996, is a cooperative body that promotes common activities and supervises the libraries' interests. The Council monitors developments within the library sector, creates new initiatives and improves cooperation among libraries as well as between libraries and third parties.

Cooperation among the polytechnic and research institute libraries, as well as public libraries, has been organized more recently and can be seen as a response to the demands of the changing operational environment. The forums coordinating activities within the university, special library and public library sectors are called councils, while the polytechnic libraries have formed the AMKIT

consortium. The management groups of the councils and the AMKIT consortium act as negotiation partners for the National Library (Hormia-Poutanen 2002a).

There are three different types of library consortium in Finland. The Linnea consortium <http://www.lib.helsinki.fi/linnea2/konsortio/index_eng.htm> was formed to handle matters related to the library system at universities and at several special libraries (Jauhiainen 2000, 2004). The licensing of electronic resources and development and coordination of the national information retrieval portal, Nelli, are handled by the FinELib consortium <http://www.nationallibrary.fi/libraries/finelib/>. Universities, polytechnics, research institutes and public libraries are members of the FinELib consortium. The Linnea and FinELib consortia were formed to develop certain digital library services. The third consortium, AMKIT <http://www.amkit.fi>, was formed to process all issues related to the development of polytechnic libraries (Hormia-Poutanen 2002a).

The organization of the Finnish library network is exceptional by international standards. Each library sector can handle questions related to the entire sector, and the councils' management groups are authorised to represent the whole sector in negotiations. There is a growing need for cross-sectoral cooperation and exchange of expertise among the different disciplines.

IT centre for research, the National Library and libraries working in partnership

The Ministry of Education administers CSC <http://www.csc.fi/>, the Finnish IT Centre for Science. CSC is a non-profit company providing IT support and resources for academia. CSC provides Finland's widest selection of scientific software and databases, Finland's most powerful supercomputing environment and a very advanced research network, Funet.

The National Library of Finland became a service centre for university, polytechnic, public and special libraries in 2006. The new role was formalized in the University Act which emphasises the importance of the library's national responsibilities. The main task of the service centre is to improve public access to information (Hormia-Poutanen 2005). The Ministry of Education provides central funding to cover the costs of staff and to some extent also software and licences. The amount of central funding is approximately 5.5 million euros annually. The services currently provided by the National Library for the library network are described in more detail in Table 1.

In 2008, the Ministry of Education sent out a clear message – the role of the National Library as a service provider should be further strengthened.

For academia, the CSC and the National Library are two particularly important major service centres. These two organisations have a long tradition of cooperation, particularly regarding library hardware maintenance. The cooperation has resulted in significant centralisation of servers running library software (Table 2). Perhaps the most important end result has been the very clear division of work between the CSC, the National Library and the individual libraries (Figure 3).

The CSC also plays an important role in the Finnish Digital Library project.

Figure 3 Division of labour between the IT center for research, the national library and libraries

Table 2 Steps taken to centralise library system hardware maintenance in Finland

Year	*Systems involved*	*Level of cooperation and centralisation*	*Funding*
1989–1993	VTLS library system implemented for universities, Parliament, Statistics Finland and the Repository Library	22 organizations implemented the same system, VTLS 17 servers	Central funding from the Ministry of Education
2001	Voyager library system purchased implemented for Linnea2 consortium (universities and major special libraries)	27 organizations purchased the same system, Voyager 1 server	Own funding from the participating organizations
2002–2004	Voyager library system implemented for polytechnic consortium	57 organisations using Voyager 2 servers	Central funding from the Ministry of Education
2005-2007	MetaLib and SFX implemented for universities, polytechnics, public libraries	68 organizations purchased the same system 3 servers altogether (Voyager + MetaLib)	Central funding from the Ministry of Education
2007	Centralized server solutions to run Voyager and MetaLib + SFX software.	1 server (SUN M9000) for universities, polytechnics, public libraries.	Central funding from the Ministry of Education

References

Hormia-Poutanen K. 2002a
The National Electronic Library in Finland, FinELib – Licensing Content for Research and Learning Environments on the Basis of User Needs. pp. 234–257
In *Global Issues in 21st Century Research Librarianship*. NORDINFOs 25th Anniversary Publication, edited by Dr Sigrún Klara Hannesdóttir. Helsinki.

Hormia-Poutanen K. 2002 b
The National Electronic Library Programme/the backbone of the information supply in the electronic research environment in Finland
Serials. *The journal for the serials community* **15**(3): pp. 219–223

Hormia-Poutanen K. 2003
Ensuring the future success of FinELib, the Finnish Electronic Library
Helsinki University Library Bulletin 2003:17–19

Hormia-Poutanen K. 2005
The National Library – providing new services for the information society
Helsinki University Library Bulletin 2005: 4–8

Hormia-Poutanen K. 2006
Consortia in Europe – describing the various solutions through four country examples
Library Trends **54**(3): 359–381

Hormia-Poutanen K. 2007
Improving access to information – defining core electronic resources for research and wellbeing
Liber Quarterly **17**(3/4)
Details available at <http://liber.library.uu.nl/>

Hormia-Poutanen K. 2008
Digital library infrastructure for libraries – challenges for the future
The National Library of Finland Bulletin 2008 <http://www.kansalliskirjasto.fi/extra/bulletin/article3.html>

Jauhiainen A. 2000
A New Library System for Finnish Research Libraries Chosen
Helsinki University Library Bulletin (12–19)

Jauhiainen A. 2004
Linda: Union Catalogue for Finnish Academic and Research Libraries, pp. 101–122
In *Union Catalogues at the Crossroad.* Edited by Andrew Lass and Richard E. Quant

Ministry of Education. 2008
National-Level Research Infrastructures: Present State and Roadmap. Summary and recommendations
Details available at <http://www.tsv.fi/tik/kansallisen_tason_tutkimusinfrastruktuurit_tiivistelma_englanti.pdf>

Information Behaviour of the Researcher of the Future. 2008
CIBER (Centre for Information Behaviour and the Evaluation of Research), UCL
Details available at <http://www.bl.uk/news/pdf/googlegen.pdf>

Varis T and Saari S. 2003
Knowledge Society in Progress. Evaluation of the Finnish Electronic Library – FinELib
Helsinki: Edita
Details available at <http://www.kka.fi/julkaisut.lasso, 4:2003>

Websites

Council for Finnish University Libraries <http://www.nationallibrary.fi/libraries/council.html>

Aalto University <http://www.aaltoyliopisto.info/en/>

AMKIT consortium <http://www.amkit.fi>

The Council of University Libraries <http://www.nationallibrary.fi/libraries/council.html>

FinELib consortium <http://www.nationallibrary.fi/libraries/finelib/>

Linnea consortium <http://www.lib.helsinki.fi/linnea2/konsortio/index_eng.htm>

CSC, IT Center for Science <http://www.csc.fi/>

The Semantic Web: past and future

Kim H Veltman
Scientific Director, Virtual Maastricht McLuhan Institute, Maastricht, NL

World Digital Libraries 2(1): 33–49

Abstract

Work on semantic networks predates the introduction of the Internet in 1968. The idea of making a semantic web on the basis of RDF (Resource Description Framework) and the World Wide Web was broached by Tim Berners Lee at WWW7 (Brisbane, 1998) and further described in a roadmap (September 1998). Since then it has become a mainstream vision with progress on many fronts. A key element of the semantic web vision that makes assertions on the Web machine-readable is a RDF. This has transformed the Internet into a Web for sharing knowledge and prepares the way for a transaction web, fundamental for business dimensions of the web by providing validation of nodes and links in the system.

In all this, the term 'semantic' is used in a very specific way: to understand the meaning of (programming) instructions to computers without human intervention. This marks an enormous contribution. Even so, the focus on 'semantic' instructions for machines, does not yet address the needs of those in the Social Sciences, the Humanities, and indeed all Sciences, which have temporal-spatial dimensions. The present approach ignores the traditional meaning of semantics, whereby it was linked with etymology: the history of meanings given to words by humans, which change with time and space. It also uses a very limited form of hypertext that links one hyperlinked word in one document with another hyperlinked word in another document.

This paper reviews briefly the history of these efforts and outlines the potentials of a semantic web with multiple levels of hyperlinks. It suggests that current efforts focus on a Web that privileges born digital materials. As the ITU has suggested, a next generation needs to include an Internet of things. We need much more. Scholars in earlier cultures distinguished between different worlds ranging from metaphysical and mental to physical, man-made, social, and creative. These ideas need to be integrated into our plans for a semantic web such that we can search for knowledge and information at different levels.

E-mail kim_veltman@hotmail.com

Introduction

Semantics as the science of meaning of words is a relatively recent term that goes back to late 19th century. One of the first English language books on semantics appeared in 1900.[1] Semantic networks became a significant topic in the computer science community in late 1960s, just prior to the advent of the Internet.[2] A seminal article by Woods (1975) foresaw how 'semantic networks, in which nodes are connected to other nodes by relationships called links, are widely used to represent knowledge and to support various algorithms that operate on that knowledge.'[3] In the following year (1976), Peter Chen established the idea of Entity Relationship diagrams.[4] The following decades saw the rise of 'is a' and 'has a'[5] as paradigms in object oriented programming and in the World Wide Web. While seemingly novel, these were simply the divisio (taxonomy) and partitio (partonomy) of classical logic in a new guise.

The idea of using HTML (HyperText Markup Language) and HTTP (HyperText Transfer Protocol) at CERN (1989/90) transformed the Internet into a www (World Wide Web). A first WWW conference took place in May 1994 (CERN, Geneva).[6] At WWW7 in April 1998 (Brisbane), Tim Berners Lee described the need for evolvability[7] and explored retrospective documentation of equivalence. This led to a Semantic Web roadmap in September 1998.[8] At WWW8 in May 1999 (Toronto) he outlined his vision of a Semantic Web,[9] beginning with a quest for 'human communication through shared knowledge' and a principle that 'anything can refer to anything'. A first conference followed in 2001.[10] Since then the semantic web has become a mainstream effort with projects all over the world.

The RDF (Resource Description Framework), that makes materials on the web machine readable, has opened immense new amounts of human communication and prepares the way for a transaction web, which is

[1] Michel Jules Alfred Bréal, *Semantics: Studies in the Science of Meaning*, London: Heinemann, 1900. This was a translation from the French. For a discussion of the historical context see the author's: 'Towards a Semantic Web for Culture,' *JoDI* (*Journal of Digital Information*), Volume 4, Issue 4, Article No. 255, 2004-03-15. Special issue on New Applications of Knowledge Organization Systems. Details available at <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/>.

[2] The Internet began in Britain (1968) and one year later in the United States (1969). M Ross Quilian, 'Word concepts: a theory and simulation of some basic semantic capabilities', *Behavioral Science* 12, 5 (Sept. 1967), San Diego, 410–430. Cf. P I Breslaw, *Experiments with a semantic network,* Edinburgh: Edinburgh University, Department of Machine Intelligence and Perception, 1969; R F Simmons; Jonathan Slocum, *Generating English discourse from semantic networks,* Austin: University of Texas at Austin, Department of Computer Sciences, 1970; David E Rumelhart, Donald A Norman, *Active semantic networks as a model of human memory*, San Diego, Calif.: University of California, Center for Human Information Processing, 1973; Carl Wilhelm Welin, *Semantic networks and case grammar*, [S.l.]: University of Stockholm, 1975.

3 Details available at <http://www.encyclopedia.com/doc/1O27-wallah.html>; William A Woods, 'What's in a Link'. Foundations for semantic Networks, In: *Representation and Understanding. Studies in Cognitive Science,* edited by D G Bobrow and A M Collins, Academic Press. Republished in Brachman and Levesque. Cf. Bill Woods, External Home Page: <http://research.sun.com/people/wwoods/>.

[4] Entity Relationship model: <http://en.wikipedia.org/wiki/Entity-relationship_model>.

[5] The public side of the library world associates these with Broader Term and Narrower Term.

[6] Details available at <http://www94.web.cern.ch/WWW94>.

[7] Tim Berners Lee, 'Evolvability': <http://www.w3.org/Talks/1998/0415-Evolvability/overview.htm>.

[8] Tim Berners Lee, 'Semantic web Roadmap': <http://www.w3.org/DesignIssues/Semantic.html>.

[9] Tim Berners Lee, 'Semantic Web': <http://www.w3.org/Talks/1999/05/www8-tbl/slide13-0.html>.

[10] *The emerging Semantic Web: selected papers from the first Semantic Web Working Symposium,* edited by Isabel Cruz, *et al.*: Amsterdam: Oxford: IOS, c2002. This has led to an annual International Semantic Web Conference. Cf. <http://iswc.semanticweb.org/>.

fundamental for business dimensions of the web. These remarkable efforts entail the greatest technological success stories of all time. It took a century to reach 100 million telephone users. While the Internet took twenty years (1969–1989) to reach 1 million users, the WWW has in the past 18 years grown by a further 1462 billion fixed-line users, with predictions of a doubling of these figures in the next few years as mobile Internet becomes the norm.[11] With a billion new mobile phones annually, it is technologically possible to foresee a world where everyone could one day be connected such that all knowledge could theoretically be shared online.

Dreams of automatic systems

Underlying the RDF is a quest for automatic verification of the nodes in web hyperlinks. The mechanical version of this quest has its roots in Antiquity with traditions of building automata.[12] Leibniz (1646–1716)[13] took this vision a decisive step forward with his 'wonderful idea of… a special language built with an alphabet of concepts, plus the algebra of logic to determine the truth-value (true or false) of sentences in this language.'[14] This led him to a *calculus ratiocinator*[15] and efforts to determine *characteristica universalis.*[16] Leibniz explored what has since been called the Symbol Grounding Problem;[17] explored automated reasoning and 'dreamt of building a system where the truth of any assertion could be determined by calculation.'[18] These were early premonitions of one version of a semantic Web.[19]

During the 1920s and 1930s, scientists in Germany, Britain, Russia and the United States prepared the way for programmable computers. In Britain, Alan Turing worked on an Automatic Computing Engine.[20] In the United States, this led to the work of Vannevar Bush[21] and his student Claude Shannon.[22] During the 1940s, these efforts took a curious turn. The quest for a programmable computer became a quest to produce M2M (machine-to-machine) communications. John von Neumann[23] became a champion of this vision that

[11] Internet Usage Statistics: <http://www.internetworldstats.com/stats.htm>.

[12] Automaton <http://en.wikipedia.org/wiki/Automaton>.

[13] Leibniz <http://en.wikipedia.org/wiki/Gottfried_Leibniz>.

[14] Prelude <http://www.cse.yorku.ca/course_archive/2006-07/F/2001/handouts/lect00.pdf>

[15] Calculus ratiocinator <http://en.wikipedia.org/wiki/Calculus_ratiocinator>.

[16] Characteristica universalis <http://en.wikipedia.org/wiki/Characteristica_universalis>.

[17] Consciousness Studies/The Philosophical Problem/Machine Consciousness <http://en.wikibooks.org/wiki/Consciousness_Studies/The_Philosophical_Problem/Machine_Consciousness>. In more recent literature these problems have become linked with John Searle's Chinese Room Argument: http://www.iep.utm.edu/c/chineser.htm .For a another discussion of these philosophical dimensions: Harry Halpin, "Identity, Reference, and Meaning on the Web", *WWW 2006, May 23-26, 2006,* Edinburgh, Scotland. ACM 1-59593-323-9/06/0005: <http://www.ibiblio.org/hhalpin/irw2006/hhalpin.html>.

[18] Automated Reasoning: <http://users.jyu.fi/~antkaij/opetus/ap/2004/aure-ch1.pdf>.

[19] Other premonitions include Vladimir Odoevsky, whose novel, Year 4338 (written 1837) outlined features pointing to a Web (http://en.wikipedia.org/wiki/Runet). Paul Otlet outlined an amazingly detailed vision in his Monde (1934) (http://www.laetusinpraesens.org/docs/otlethyp.php). Also astounding was the Russian David Sarnoff (1964) (http://earlyradiohistory.us/).

[20] B Jack Copeland, Alan Turing's Automatic Computing Engine. The Master Codebreaker's Struggle to build the Modern Computer, Oxford: Oxford University Press, 2005.

[21] Vannevar Bush: <http://en.wikipedia.org/wiki/Vannevar_Bush>.

[22] Claude E Shannon, *The Mathematical Theory of Communication,* , Illinois: University of Illinois Press, 1949. The preface was written by Warren Weaver, also head of the scientific section of the Rockefeller Institute.

[23] John von Neumann, *Papers*, Library of Congress, Box 33: <http://www.loc.gov/rr/mss/text/vonneumn.html>: Reports 1946, 'Preliminary Discussion of the Logical Design of an Electronic Computing Instrument'. 1948, 'Planning and Coding of Problems for an Electronic Computing Instrument' (3 folders).

machine-to-machine communication might one day entirely replace any role for humans in the decision process.

Norbert Wiener[24] saw the dangers of such an approach and argued for an alternative vision, which kept humans an integral part of the automation process. When his efforts were ignored, he developed the field of cybernetics (1948)[25] and subsequently inspired the first cybernetics institute in Naples (1968).[26] The dangers of machine-to-machine communication were further explored by Joseph Weizenbaum (1976)[27] and Grant Fjermedal (1986)[28] and at a popular level in films such as War Games (1983),[29] where a supercomputer named WOPR (War Operations Programmed Response)[30] almost destroyed the world because it made no distinction between simulated game and reality.

Such dangers are far removed from the ideals of the W3 Consortium. History should remind us, however, that the intentions of inventors are often far removed from those who apply these inventions. Parallel with the military, moreover, is a civilian dimension of M2M as a new kind of pervasive internet, where the market for M2M devices is expected to reach $300 billion by 2010.[31] Such interests could lead to very different directions.

As long as the Web remains a place where everyone is free to express their thoughts, within the limits of decency and basic legal frameworks, these dangers would appear to be minimal. But what if standard search engines, or extremely large consortia developed PVNS (Private Virtual Networks) and Intranets with different standards and possibly agendas contrary to the public good? Or even, as has occurred in the past, totalitarian regimes gain supremacy? What if they took control of sections of the web, perhaps even of whole countries? This could lead to forms of censorship where we have no simple way of determining (a) what is being omitted or (b) which is being presented is true or false. For such eventualities, we need more than the present approach to verification and truth.

Verification and truth

The semantic web in its present form has proof validation but not proof generation.[32] It strives to let everyone express their notions of truth, but is very skeptical of deciding about truth.[33] It is useful to quote at length what the Web strives to achieve.

[24] Norbert Wiener: <http://en.wikipedia.org/wiki/Norbert_Wiener>.

[25] Norbert Wiener, *Cybernetics: Or the Control and Communication in the Animal and the Machine.* Cambridge, MA: MIT Press, 1948. Ibid., T*he Human Use of Human Beings.* New York: Da Capo Press, 1950. The differences between Neumann and Wiener have been explored by Steve J Heims, *John von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death.* MIT Press, 1980. Possibly because Wiener refused to become entangled in what has been called the military industrial complex there has been an ongoing strand of criticism of everything connected with cybernetics in general and Wiener in particular. For a recent example, see: Creighton Cody Jones, 'How Wiener Attempted to Kill Science', *Executive Intelligence Review*, 4 January 2008 issue <http://www.larouchepub.com/lym/2008/3501wiener_killed_sci.html>.

[26] L'Istituto di Cibernetica 'E. Caianiello' (ICIB) <http://perseo.cib.na.cnr.it/cibcnr/chisiamo/. The roots of the laboratory go back to a seminar by Wiener at the University of Rome (1954) attended by Caianello. See: http://www.iiassvietri.it/caianiel.html>.

[27] Joseph Weizenbaum, Computer Power and Human Reason: From Judgment To Computation, San Francisco: W H Freeman, 1976.

[28] Grant Fjermedal, *The Tomorrow Makers: a Brave New World of Living Brain Machines*, Redmond: Tempus Books, 1986.

[29] War Games <http://en.wikipedia.org/wiki/WarGames. *Space Odyssey* 2001 (1968), *Demon Seed* (1977) *and I Robot* (2004) are other films that explore the theme of computers taking a path independent of their 'owners'. This is leading to a genre called Cybernetic revolt: <http://en.wikipedia.org/wiki/Cybernetic_revolt>.

[30] WOPR: <http://en.wikip WOPR edia.org/wiki/>.

[31] Machine to Machine: <http://en.wikipedia.org/wiki/Machine_to_Machine>.

[32] Tim Berners Lee, 'What the Semantic Web can represent' (also referred to as What the Web is not <http://www.w3.org/DesignIssues/RDFnot.html>.

> The Web works th[r]ough anyone being (technically) allowed to say anything about anything. This means that a relationship between two objects may be stored apart from any other information about the two objects. This is different from object-oriented systems often used to implement ER models, which generally assume that information about an object is stored in an object: the definition of the class of an object defines the storage implied for its properties.

For example, one person may define a vehicle as having a number of wheels and a weight and a length, but not foresee a colour. This will not stop another person making the assertion that a given car is red, using the colour vocabulary from elsewhere.

Apart from this simple but significant change, many concepts involved in the ER modelling take across directly onto the Semantic Web model.[34]

The principle that everyone is free to make any assertions they like is fundamental to a democratic way of life. Even so, some things are true, while others are not. If there is a yellow car in my parking lot, no amount of assertions that the car is red, blue or some other colour will change the yellow colour of the car in my parking lot in the way that a paint job could. Hence, the Semantic Web may (a) make assertions machine readable; (b) tell us whether a passage is in proper XML format, that is, whether the instructions are in correct form; (c) validate that the nodes of the triples connecting to URIs are correct, but this still tells us nothing about the meaning and ultimately the truthfulness of what is being asserted. It only confirms that the system transporting those assertions is correct.

In the case of assertions about whether a car is yellow or red, this may not seem very important. But if we want to buy a car and have an aversion to yellow, then the actual colour of the car becomes immediately significant. Or if the claim is about the mechanical condition of the car we are wanting to buy through an online site, accurate assertions about its state are vital in order to make a wise investment. At the moment, the framework focuses more on the truth of the links and connections between documents rather than the veracity of the contents of documents.[35]

Assertions and scholarship

The W3's 'principal principle' that 'anything can refer to anything' is very appealing as a manifest of freedom of thought. It corresponds to our notions of what an author of fiction should be able to do. However, this is very far removed from the traditional goals of scholarship. When a scholar makes an assertion he/she is expected to provide a reference to the source. If we doubt the claims made by an author we can return to the original source to check the author's accuracy, interpretation, and possibly arrive at a different conclusion. Very simply a scholar can refer to anything and claim anything as long as there is a link back to the objects or sources about which the assertion or claim is being made.

In past centuries, these references have typically taken the form of footnotes or endnotes, that is, they remained within the confines of an author's article or book. A limitation of this approach was that in the case of rare sources, tracking down the original in

[33] Tim Berners-Lee, 'The World Wide Web and the "Web of Life," ' with a paragraph on Truth <http://www.w3.org/People/Berners-Lee/UU.html>.

[34] As in note 30: Tim Berners Lee, 'What the Semantic Web can represent': <http://www.w3.org/DesignIssues/RDFnot.html>.

[35] Some would argue that there are sometimes tensions between the rhetoric of letting everyone say anything about anything and the quest to have us create triples for everything we want to express.

some location beyond the author's publication was often a difficult task. In an electronic environment, such references in a publication can potentially be linked directly with objects outside an author's text. To take a simple example, assertions that the Mona Lisa by Leonardo da Vinci is in the Louvre Museum concern an object outside the Web. To test its veracity requires checking whether the Louvre has such a painting and certifying that it is indeed still at that location. One solution would be to insert an RFID sensor into every painting including the Mona Lisa and to determine whether the painting is in the place as asserted. Alternatively, a video camera could confirm that the painting is still in the section, Denon, First Floor, Salle des Etats.[36] In this case, a hyperlink would not only show us the original in situ but also give us an indication of its present state.[37] Proving that the video link is exact and that it has not been spoofed remains elusive however. There is no simple technological fix to the quest for truth but that does not diminish the importance of the quest.

Powers and limits of Born Digital

The near trivial examples cited above of a yellow car in a parking lot and a painting in the Louvre point to a fundamental limitation of the Semantic Web as it exists today. It is about hyperlinks within a www as a self-enclosed system operating through an Internet of cables, wires and wireless connections. This is partly because the quest for a Semantic Web is focused on the verification of URI (Uniform Resource Identifier) nodes, RDF triples and XML syntax as if the Web in isolation were enough. The power of this approach is that the system can determine whether the statements within the Web are logically correct. This has also contributed to a rise in attention to so-called Born Digital information.

In its present form, the Resource Description Framework focuses on the correctness of descriptions of resources, that is, on the logical correctness of annotations and commentaries about sources, rather than providing a method for linking claims with original sources, or for checking the veracity of statements about objects that are not digitally born. Paintings in the Louvre are but one example. The objects discussed and about which assertions are made on the WWW are increasingly in multiple 'worlds' beyond the WWW. Modern astronomy is providing us with vast quantities of information from space. Projects such as Google Earth and Microsoft Live Local are creating a one to one model of the earth. Similar projects are underway for the oceans. To make scholarly assertions about these materials of the heavens, earth or the oceans requires linking back to the original 'images' and objects on

[36] Leonardo, *Mona Lisa*, Louvre: <http://www.louvre.fr/llv/activite_detail_parcours.jsp?CURRENT_LLV_PARCOURS%3C%3Ecnt_id=10134198674098115&CURRENT_LLV_CHEMINEMENT%3C%3Ecnt_id=10134198674098123&CONTENT%3C%3Ecnt_id=10134198674098123&bmLocale=en>.

[37] Robert Kahn, the inventor of TCP, has suggested that Digital Object Identifiers (DOIs) could be used to tag every object and living person (personal communication, ISOC Meeting, Arlington, June 2001); cf. <http://en.wikipedia.org/wiki/Bob_Kahn>. Others predict that Radio Frequency Identification (RFID) could be used for such tagging. RFID is already being used in inventory control, in animals and in some human cases. The use of RFID for Human Identity Verification remains a subject of considerable debate. The same technologies being sponsored as tools for helping protect our identity also open new avenues of identity theft.. There is clearly a delicate balance between the need to track constantly rare individuals who are truly a threat to society and the equally great need to assure the continued privacy and non-tracking of a majority of citizens. Cf. Data Privacy & Integrity Advisory Committee. The Use of RFID for Human Identify Verification, Report No. 2006-02: No. 2006-02: <http://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_12-2006_rpt_RFID.pdf; cf. RFID….Controversies: http://en.wikipedia.org/wiki/RFID#Controversies. Some see (RFID) as a key part of a global surveillance society: RFID: http://www.oilempire.us/rfid.html>.

which our claims are based. In the case of a distant galaxy, a born digital document may be all that it is feasible to have. In many cases, however, we need to be able to link back to the original objects, materials or sources.

To do so effectively, it is important to introduce traditional distinctions into the digital environment. Some of the sources are physical (trees, plants, animals). Some are man-made (houses, buildings, churches, synagogues, mosques, temples). Some are social. Some are mental (literature). Some are metaphysical (metaphysics, religion, mythology). We need tools that can distinguish these. A future version of a semantic web should allow us not only to search for butterflies, but also allow us to specify whether we are concerned with physical specimens, artistic images, literary versions, man-made examples or social butterflies.

This could at first seem an unreasonable goal until we recall that the discipline based approach to classification in libraries and memory institutions has already achieved such distinctions. Knowledge concerning physical specimens of butterflies is found under biology. Artistic images of butterflies are found under art. Literary examples of butterflies are found under literature, and so on. There is no need to re-invent the wheel. The challenge lies in integrating existing classifications of memory institutions (as a kind of metadata) into the filtering processes of search engines. Instead, of receiving a list of millions of undifferentiated hits we would get a list that shows us how many butterflies are found under biology, art, literature, and so on. If these lists are then subdivided by time (when), place (where), conditions (how), causes, motivations (why) as well as the usual who and what, current lists which are too long will become usable.

Current limits of RDF

To achieve this, the present RDF and its approach using triples are not sufficient. RDF currently has a list of media descriptors (screen, tty, tv, projection, handheld, print, braille, aural, all),[38] which provide a useful list of media display formats, that is, the output side of the process. Meanwhile, IANA's MIME types, based on RFC 2046 (on Media Types), addresses the problem of media inputs,[39] but these are general rather than specific. Hence, they distinguish text as a basic media type, but do not tell us whether it is a book, article, manuscript, and so on. (cf. material selection and formats in the library world). A Resource Description Framework in a larger sense needs to provide detailed references about sources, which are ultimately the foundations of any scholarly claim.

RDF is based on model theory, which 'tries to be metaphysically and ontologically neutral'. Its basic idea is appealing.

The idea is to provide an abstract, mathematical account of the properties that any such interpretation must have, making as few assumptions as possible about its actual nature or intrinsic structure, thereby retaining as much generality as possible. The chief utility of a formal semantic theory is not to provide any deep analysis of the nature of the things being described by the language or to suggest any particular processing model, but rather to provide a technical way to determine when inference processes are valid, that is, when they preserve truth. This provides the maximal freedom for implementations while preserving a globally coherent notion of meaning.[40]

[38] Media Descriptors: http://www.w3.org/TR/html401/types.html#h-6.13

[39] RFC 2046: http://www.ietf.org/rfc/rfc2046.txt. See also IANA Mime Types: http://www.iana.org/assignments/media-types/

[40] RDF: http://www.w3.org/TR/2004/REC-rdf-mt-20040210/

This is excellent for universally[41] true assertions/statements about who and what (is a and has a, also called taxonomy and partonomy). The authors explain that RDF 'does not provide any analysis of time-varying data'. As a result, the quest for simple statements in RDF means that statements about entities (what) are privileged over possible statements about where, when, how and why.

The stated goal is to preserve truth. Paradoxically the 'truth' of living beings such as humans requires that we descend from abstract concepts to specific, individual cases with a given time (when), place (where), under given conditions, using particular means (how) and usually with given motivations, beliefs, reasons, causes (why). Thus, the quest to create a machine readable code produces text (a) unreadable by an average person, which (b) does not capture the specificities of persons. Hence, a Web that theoretically aims at 'anyone being (technically) allowed to say anything about anything,'[42] in practice requires that we limit ourselves to a small sample of possible relations (cf. Appendix 1).[43] Speaking poetically, we gain the black and white of logic and lose the rainbow colours of life. This is a long way from the universal promises made by some zealots in the field.

New knowledge	Science, new E-Science
Information, transactions	Business, E-Business
Creative expression and informal knowledge	Literature, Poetry, Art, new E-Culture
Enduring knowledge	Memory institutions, E-Culture, E-Science

Figure 1 Four goals for the Web

Triples

These limitations are built into the triples 'architecture' which unites (a) a subject via (b) a predicate (also called a property) to (c) an object. This is adequate for an assertion such as 'Gazza Ladra is a thief' but inadequate to state precisely that: 'Gazza Ladra stole the jewel from Madame Castafiore at three p.m. on Thursday at Marlinspike by hiding it in a bird's nest.' And real life robberies are often more complex than those in Tintin.

The authors of the text on RDF semantics may speak of worlds and universes as if they admitted multiple interpretations over a period of time but ultimately the 'architecture' of a system based on abstract universals assumes a static reality that does not change over a time: a 'meaning' that is as constant as the laws of mathematics. This may well be what is needed for the framework. However, those in the social sciences and indeed all the temporal sciences (for example, geology, geography), have very different needs.

Partly because human beings are affected by time, place and conditions, the meanings that they assign to terms change over a period of time. That is why we have etymological dictionaries and disciplines such as philology and historical linguistics. Modern authors of fiction may want complete freedom in

[41] Whether such statements are indeed universally true depends of course on whether the person making the 'is a' and 'has a' statements has created an accurate classification in terms of taxonomy and partonomy. To illustrate the validity of their method, computer scientists invariably use an obvious case such as 'John Doe is a man.' But as anyone who has worked in taxonomy or the profession of indexing knows, there are many cases when classing is not an obvious task. When carried out by non-professionals in indexing, the claims for universal truth of computer scientists are often far more shaky than their rhetoric would have us believe.

[42] As in note 30: Tim Berners Lee, 'What the Semantic Web can represent': <http://www.w3.org/DesignIssues/RDFnot.html>

[43] '*Contents-related relations* are likewise fourfold and can be grouped into 1) generic, 2) partitive, 3) opposition/complementarity, and 4) functional relations.' Dahlberg (unpublished manuscript, 2008). Cf. the fundamental study by I. Dahlberg, *Grundlagen universaler Wissensordnung. Probleme und Möglichkeiten eines universalen Klassifikationssystems des Wissens*. München: K.G.Saur Verlag, 1974.

annotating as they please. Meanwhile, historians have a very different task of discovering the meanings of words used by individuals in different times and places. To achieve this, well defined RDF syntax and semantics could help us link to historical texts or sources. But we need much more if we wish to understand the meaning of those sources. The commitment to triples (or truples) needs to be expanded at least to sextuples.

New scale and scope of the Web

In all fairness, it is important to acknowledge that the unexpected success of the WWW has completely transformed the scale, and scope of the challenges it faces. Moreover, accelerating, evolutionary, adaptability of emerging technology convergence is about to multiply the dimensions and layers of this scale and these challenges. At the outset in 1989/90, the WWW was effectively trying to solve problems that faced the Internet: sharing knowledge between/among small communities of researchers in high energy physics, astronomy, chemistry, and advanced researchers in other fields. This amounted to less than a million persons. Today the W3C is faced with a web of over 1.6 billion users with approximately 20 million new users per month. In 1990, 'content' was mainly in the form of e-mails and pre-prints of high level scientific papers. By 2000, an estimated 7 million pages of materials were being added daily. There are now plans, which, by 2020, the full contents of over 60 million books and documents will be available online.[44] If the Internet began as a prototype for what is now seen as E-Science, the WWW needs to become a prototype for both E-Science and E-Culture.

Apart from obvious shifts in scale, this transforms the kinds of relations that are involved. Those in computer science typically speak of two kinds of relations: 'is a' and 'has a', that is, taxonomy and partonomy.[45] Meanwhile, those in the field of knowledge organization speak of at least 12 basic kinds of relations (Appendix 1).[46] We need a wider vision that integrates all these relations and also brings to light their historical evolution.

Four goals

The new scale and scope of the Web has also shifted the potential goals of the Web. In the first decade of the 20th century, Paul Otlet, Henri Lafontaine and a group of thinkers developed a vision of a World Brain (Gehirn der Welt), which foresaw making the whole of enduring knowledge accessible online. They also foresaw what was later called hypertext and hypermedia.[47] The American Internet (1969–) reduced this to a much narrower goal: new knowledge in the form of science (now e-science). Pioneers such as Andreas Okopenko

[44] For a fuller discussion see the author's: "Framework for Long-term Digital Preservation from Political and Scientific Viewpoints: Rahmenbedingungen der digitalen Langzeitarchivierung aus politischer und wissenschaftlicher Sicht, '*Digitale Langzeitarchivierung. Strategien und Praxis europäischer Kooperation, Deutschen Nationalbibliothek, anlässlich der EU-Ratspräsidentschaft Deutschlands, 20-21. April 2007*, Frankfurt: National Bibliothek, 2007 (In Press). http://www.sumscorp.com/articles/pdf/2007_From_Recorded_World_to_Recording_ Worlds.pdf.

[45] While logic is constantly mentioned, not everyone in the field of computer science is always clear about the precise differences even between *divisio* and *partitio*. Woods (1991) saw subsumption and taxonomy as a basis for a new framework. William A Woods, 'Understanding Subsumption and Taxonomy: A Framework for Progress,' in John Sowa (ed.), *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, San Mateo: Morgan Kaufmann, 1991.

[46] Categorial relations are directly linked with the definitions of words, terms and concepts. So a first challenge lies in linking every word we use with its various meanings. This need is all the more pressing since those in recent fields such as computer science, and cognitive science insist on using familiar terms with a long history of meaning in completely different ways such that true interdisciplinary discussions are increasingly difficult.

[47] Anthony Judge, 'Paul Otlet's 100-year Hypertext Conundrum ?', laetus inpraesens, 28 May 2001, Brussels: UIA: http://www.laetusinpraesens.org/docs/otlethyp.php.

in Europe and Ted Nelson in the US saw these developments as a new key to creative expression (literature, poetry), although visionaries such as McLuhan also warned of the dangers of electric and electronic communication.

The advent of the WWW in the 1990s introduced new possibilities for both (a) new knowledge and (b) creative expression and initially the quest for 'anyone being (technically) allowed to say anything about anything', seemed an obvious and unassailable goal. Work on software systems such as Annotea seemed to support this direction. However, this widening of possibilities also posed a need to distinguish (a) verified knowledge from (b) creative, informal knowledge. To answer this need the W3C began by separating reason, (which could be tested by logical proof) from rhyme, which could not. Next they sought to make reason not only logical but also machine readable. This had the great advantage that it seemed to make the goals of science for new knowledge the same as the goals of business for accurate information about transactions. As this became clear in the years 1997–1999, the WWW gained the support of big business. The quests for E-Science and the E-Business seemed complementary. Both wanted a way of verifying knowledge and information.

While the rhetoric of allowing anyone to say anything about anything continued, the W3C's work on annotation software slipped into the background; as did the creative strand symbolized by figures such as Ted Nelson. The dot-com bust (2000/2001) seemed to challenge the E-Business vision, but merely eliminated companies without clear business plans. Meanwhile, the past eight years have seen an amazing rise of an international Web. In 1988, over 90% of the Internet was in the US. In 2008, the US represents less than 17% of the Web. English now represents less than 30% of the Web. If the Internet initially focused on one goal of new knowledge, the WWW now needs to address four goals: (a) new knowledge of research; (b) information of transactions; (c) enduring knowledge of memory institutions and (d) informal knowledge and information of everyday users ranging from the frivolity of an SMS to the heated discussions, philosophical and otherwise of blogs, chats, and so on. (Figure 1).

One fundamental consequence[48] is that the quest for E- Science needs to be complemented by a quest for E-Culture. A combination of E-Science and E-Culture, requires that we keep intact the results of persons who have said/written things about things (and about persons) in different ways at different times. This cannot be achieved by having only one type of knowledge representation. Hence, a deeper challenge is to maintain alternative relations in knowledge structures, including some which are no longer fashionable. We need to acknowledge and sustain what Francis Bacon coined different 'knowledge' as different ways of thinking, different ways of ordering the world, else the Web risks imposing a colonialism of bits, which is as threatening as the earlier forms.

Evolvability on a new scale

When Tim Berners Lee spoke of evolvability and retrospective documentation of equivalence in 1998,[49] few suspected that this would need to apply to over five thousand years of documentation but, in a world where

[48] Another consequence is that, while the semantic web strives for a single framework, the four goals imply four different sets of criteria for the Web. Just as we already distinguish between .com, .org etc., we need to distinguish accept that some goals require verification while others do not.

[49] A decade ago, some individuals in the United States assumed that they were destined to decide how this might be done. Today, in a world where the U.S. represents less than 17% of world users, this is no longer obvious, especially if the same country wishes to uphold claims for democratic decisions by majorities.

(a) the full contents of 60 million books are planned and (b) there are plans for 50% of the world's population to be online within seven years, these are effectively the new system requirements. We need a new architecture that also has retrospective documentation of equivalence and non-equivalence, different ways of ordering, classifying and understanding the world.

Multiple level links

As the Internet was being planned in the 1960s, pioneers such as Ted Nelson[50] and Andreas Okopenko,[51] explored the potentials of hypertext and hypermedia. Partly due to a narrowing of goals described above, hypertext on the Web has become a much more limited concept: a hyperlink on the Web typically links us to only one other location. Ted Nelson, who coined the term 'hypertext', distances himself from these recent developments.[52] There is even a 'dispute over whether the Web is a hypertext system at all.'[53]

[50] Ted Nelson <http://ted.hyperland.com/>.

[51] Ted Nelson has been credited with coining the term (1965) but the first practical demonstrations were probably by Andreas Okopenko in Vienna (1968–1970). For a discussion and further references see the author's *Understanding New Media* under Hypertext: <http://www.sumscorp.com/kavai/newmedia/>.

[52] Ted Nelson: 'The World Wide Web (another imitation of paper) trivializes our original hypertext model with one-way ever-breaking links and no management of version or contents.' <http://www.xanadu.com/xuhome.html>.

[53] Robert Cailliau, Helen Ashman, 'Hypertext in the Web—a History', *ACM Computing Surveys* **31**(4), December 1999. <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/62.html>. Cailliau cites: [Nürnberg 1999] Peter J. Nürnberg and Helen Ashman. 'What was the Question? Reconciling Open Hypermedia and World Wide Web Research' in Proceedings of ACM Hypertext '99, Darmstadt, Germany, 83–90, February 1999. For one discussion of other possibilities: Douglas Tudhope and Daniel Cunliffe, 'Semantically Indexed Hypermedia: Linking Information Disciplines', *ACM Computing Surveys* 31(4), December 1999: <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/6.html>.

The same HTML, XML and RDF, which are now being used to create only one set of one directional links, can be adapted to achieve much more. For instance, a team of young Russians, headed by Vasily and Alexander Churanov (Smolensk) have demonstrated the principle of omni-links, whereby every word in a text is (a) hyperlinked and (b) linked at multiple levels.[54] These principles are being further developed in a site on New Models of Culture.[55] Here a hyperlinked word can lead to (a) definitions in dictionaries,

Traditional concordances distinguished between Key Words In Context (KWIC) and Key Words Out of Context (KWOC). The Web has both Cross References In-Context (XRIC) and Cross References Out-of-Context (XROC), with a reported preference for XRIC. Cross-Reference: http://en.wikipedia.org/wiki/Cross-reference. Another problem with current hyperlinks is that they are typically highlighted in blue, which traditional readers often find distracting.

[54] This is available as CD Rom and online in an experimental demonstration See: http://www.sumscorp.com/kavai/newmedia_cd/ Choose: Books, then New Media, then Omnilinked. The text appears as it does in a regular publication with no highlighted words because every word is linked. So if , for instance, we choose the word Augmented in the Title we immediately have access to 49 titles relating to augmented and 14 keywords: including Augmented Objects and Experiences, which takes us to a publication on Augmented Animals. Alternatively if we choose Full Text Search in New Media we are given a list which shows how often the word Augmented was used in each chapter. Chapters with no occurrences remain not highlighted. Clicking on the highlighted chapter names takes us to those occurrences. Full text search: Augmented
1. Preface (1); 2. Introduction (1); 3. Acknowledgments; 4. Computers (3); 5. Mobility (2); 6. Miniaturization
7. Production (1); 8. Services (4); 9. Institutions (4)
10. Organizations; 11. (Knowledge) Management (1);
12. Learning (1); 13. Personal Knowledge;
14. Collaboration (5); 15. Enduring Knowledge (21)
16. Challenges; 17. Synthesis (30) 18. Conclusions (13)
19. Epilogue 1; 20. Epilogue 2 (1); 21. Appendices (1)
22. Illustrations (1); 23. Abbreviations (2)

If we go back and choose Other Definitions the system goes to One Look Dictionary, which then finds 18 definitions for Augmented in a range of online dictionaries.

[55] New Models: http://sumscorp.com/kavai/newmethods/

(b) explanations in encyclopaedias, and (c) titles of articles and books.

Such choices reflect basic realities of knowledge organization, which is at many levels range from a single term (terminology, thesaurus, classification); to a definition (dictionaries); to an explanation (encyclopaedias); to titles (library catalogues, bibliographies); to partial contents (abstracts, reviews); to full contents (full texts of articles, books, and so on). In major libraries these various levels of knowledge are collected in different sections. Dictionaries and encyclopaedias are in one part of a reference room; library catalogues, book catalogues, bibliographies, citation indexes, reviews in other parts, while periodicals and books have their own stacks and in large institutions sometimes have separate buildings.

Contemporary search engines offer us random examples of knowledge and information from all these levels. Needed is a more systematic approach that offers the electronic equivalent of a library reference room. A semantic web in this deeper sense will (a) allow us to navigate at will to the appropriate level of knowledge. So every hyperlink will have a one-to-many (levels) feature. Moreover, the links must go beyond nodes in hyperspace (URIs) and take us back to the original sources, that is, new kind of references that takes us beyond a note at the foot of a page or end of a book to go beyond the book to the original text, image, object or source being cited.

We thus envisage common interfaces[56] for a multilevel approach to basic choices (Appendix 2). Earlier work[57] has explored the potentials of Virtual Reference Rooms with up to 12 levels of knowledge. These can be seen as Knowledge Choices (K Choices, Appendix 3). There would be three modes of searching. An initial Look Up Mode divided into Basic, Intermediate and Advanced would be for regular use. A Study Mode would offer further possibilities for students. Researchers would have a Research Mode. Users with basic needs would choose from one of ten Goals Choices (G Choices, Appendix 4). Researchers would have access to Strategy Choices (S Choices, Appendix 5) which reflect both Terms Choices (T Choices, Appendix 6) and Source Choices (SO Choices, Appendix 7). In such an approach, the notion of a highlighted word of hypertext linked to one other resource would be transformed into a multilayered systematic approach.[58]

As noted earlier, the W3C, and computer science generally, have focused on 'is a' and 'has a' relations, mainly in the creation of new relations as annotations to modern documents. Scholars and researchers at centres of learning and memory institutions in particular have been making commentaries, annotations and relations for millennia. As the full-text contents of memory institutions become available online, larger challenges loom. How do we create seamless mappings between established systems and the new frameworks? How do we guarantee the authenticity of archived mapping and the veracity/and or completeness of mappings as they are presented to the 'general public'. The past decade has seen numerous useful steps in

[56] At present, every medium tends to have its own remote control. We are told that, by 2015, these various media will all be connected within one digital framework. We need a common interface for such devices. The author's home has different remote (controls) for 1) regular television, 2) digital channels 3) video, 4) DVD ; 5) computer and 6) telephone . For a more detailed discussion of these trends see the author's *Understanding New Media*, Calgary 2006.

[57] See the author's: "Access, Claims and Quality on the Internet: Future Challenges", *Progress in Informatics*, Tokyo, no. 2, November 2005, pp. 17–40. <http://www.nii.ac.jp/pi/n2/2_17.pdf.>; *Understanding New Media. Augmented Knowledge and Culture*, Calgary: University of Calgary Press, 2006.

[58] A preliminary demo is available at <http://sumscorp.com/kavai/newmethods/>. A new multilingual version is being developed. Details available at <http://195.28.20.73/>.

this direction including SKOS.[59] Even Wiki is beginning to map its categories with Dewey[60] and Library of Congress classes. Eventually we need systems that allow us to move seamlessly between different classification systems; to be able to search for knowledge and information using any of the main kinds of relations. Two, more elusive, goals are (1) an historical system that allows us to see how world views change as access to sources moved from a few hundred to thousands, millions and now potentially billions of items and (2) an emergent system that allows us (a) to visualize more clearly possible relations in new fields of knowledge and (b) recognize areas of research, which have not yet been explored or deserve further attention.

Worlds Wide Webs

The remarkable evolution of the WWW over the past two decades has been in the context of unprecedented change. While some members of the W3C have necessarily focused on a basic infrastructure, others have explored more dramatic possibilities. For instance, at the Internet Conference (Yokohama, 2000), the visionary Professor, Jun Murai, outlined a world wherein every automobile might one day have c. 140 online connections to knowledge bases. In this approach, the Web extends far beyond the networks of wires and wireless connections and links directly with the physical world. Elements of such a vision have already been incorporated in an RDF Primer (2004).

> Like HTML, this RDF/XML is machine processable and, using URIs, can link pieces of information across the Web. However, unlike conventional hypertext, RDF URIs can refer to any identifiable thing, including things that may not be directly retrievable on the Web (such as the person Eric Miller). The result is that in addition to describing such things as Web pages, RDF can also describe cars, businesses, people, news events, and so on. In addition, RDF properties themselves have URIs, to precisely identify the relationships that exist between the linked items.[61]

Describing objects in the physical world is one step. Linking to them directly via RFID chips, sensors, web cams, and other devices would mark a further step.[62] Today's WWW operates on/in the infrastructure provided by the Internet. Increasingly, the WWW is becoming intertwined with the worlds of the heavens (Space Sciences), the earth (Earth Sciences) and the oceans (Ocean Sciences) such that it makes sense to speak of Worlds Wide Webs. These worlds include metaphysical worlds which have evolved in various cultures over time, for example, lokas of the Hindus and the Buddhists and olams of the Essenes and Hebrews.

Physical world as knowledge interface

All this is transforming our sense of where various electronic and other worlds begin and end. A generation ago, the Internet was pictured as a cloud with IP addresses at the extremities to indicate terminals. Computers were mainframes, desktops or laptops. Today, they are also mobile devices linked with cameras and other tools. In the past, we studied the physical world and then stored the results of our studies in memory institutions, that is, LAM (Libraries, Archives and Museums). In the past, cameras were passive devices. This is changing. We might still use a

[59] SKOS: http://en.wikipedia.org/wiki/SKOS

[60] List of Dewey Classes: http://en.wikipedia.org/wiki/List_of_Dewey_Decimal_classes

[61] RDF Primer: http://www.w3.org/TR/REC-rdf-syntax/

[62] Radio Frequency identification: http://en.wikipedia.org/wiki/RFID. Potential dangers to privacy through implanting such chips and tracking devices in humans remain a matter of concern.

camera as if it were taking a passive image, say, of a fungus that we find on a walk, or a building we see on a tour. But then that image could be sent to a memory institution, where image recognition software scans the image; compares it with digital collections in the library and then provides us with information about that which we have just seen. In such scenarios, the old saying of 'all the world's a stage' would become: all the world's an interface and unknown environments would become a context for learning in real time.[63]

Conclusion

Although an interest in meaning and interpretation is probably as old as language itself, the study of semantics as a science of meaning began just over a century ago; the use of semantic networks in computer science began just under a half century ago, and the vision of a Semantic Web only a decade ago. In that decade, the WWW has grown from 101 million (January 1998) to 1.463 billion (June 2008).[64]

The quest for a semantic web is intimately connected with an evolving RDF, which aims at machine readable code. This framework makes an enormous contribution in fostering interchange of knowledge and information. To date, this framework has focused more on commentaries than on sources; more on resources as outputs than on resources as inputs; more on born-digital electronic resources than on the vast resources in memory institutions and in the physical world. The initial framework has focused particularly on the verification of nodes as URIs. The imperative of standardized and open archive authenticity throughout the networks is a growing need.

The quest for the generality of mathematical logic has favoured abstract, universal, categories over concrete, particular, specific, and individual ones. This is excellent for the verification of nodes, which are vital for M2M communication, but of more limited use in the social sciences and humanities, where H2H (human-to-human) communication reflects the experiences of living beings for whom time, space, conditions and motivations play an important role. It is also of limited use to time bound sciences such as geology, which study long term changes rather than eternal truths. Put simply the Semantic Web thus far has focused on entities, on what and who, more than on where, when, how and why.

In M2M, what and who are often enough and a limited form of hypertext that connects one hyperlinked word in one document with another hyperlinked word in another document is often sufficient. In H2H, all six questions (who?, what?, where?, when?, how? and why?)[65] are needed as are multilevel hyperlinks. Machines are content with DTDs (document type definitions) which, ideally, remain unaffected by time. Humans require etymological dictionary definitions, which necessarily evolve with time.

This paper has outlined the potentials of a semantic web with multiple levels of hyperlinks. Current efforts focus mainly on a Web that is closed in the sense that it remains Web-centric even if it urges open standards. As the ITU has suggested, a next generation

[63] This scenario has been further explored in the author's: 'The New Book of Nature', *eARCOM 07. Sistemi informativi per l'Architettura Convegno Internazionale,* Con il Patrocinio di UNESCO. Ministero dei Beni Culturali, CIPA, Regione Marche, Ancona-Portonovo Hotel La Fonte, 17-18-19 Maggio 2007, Ancona: Alinea Editrice, 2007, pp. 659–669.Cf: <http://www.sumscorp.com/articles/pdf/2007_New_Book_of_Nature.pdf>.

[64] Nua <http://www.dns.net/andras/stats.html#users; Internet World Stats: http://www.internetworldstats.com/stats.htm>.

[65] Dahlberg, (2008) unpublished manuscript has shown that there are at least 17 questions in knowledge organization. So even the six questions now associated with journalism via Kipling are a simplification.

needs to include an Internet of things.[66] We need much more. Scholars in earlier cultures distinguished between different worlds ranging from metaphysical and mental to physical, man-made, social, and creative. These ideas need to be integrated into our plans for a semantic web such that we can search for knowledge and information at different levels. The WWW needs to become Worlds Wide Webs. We noted that the quest for M2M communication is historically linked with a quest to remove humans entirely from the decision process. Verification of nodes is very important. A commitment to truth is even more important even if it is not a requirement for the day to day functioning of machines. We must be very careful to ensure that the steps necessary to keep the machines running, do not overshadow the larger purpose for which the machines were built: to help and foster the activities and reflections of humans.

Acknowledgements

I thank Jonathan Robin for reading the text and offering suggestions. I am very grateful to a circle of scholars and friends who continue to encourage my work. These include Madhu Acharya (Kathmandu), Professor Frederic Andres (Tokyo), Professoressa Giuseppina Saccaro del Buffa Battisti (Rome), Professoressa Francesca Bocchi (Bologna), Professor Arturo Colorado (Madrid), Professor André Corboz (Geneva), Professor Edgard Costa Oliveira (Brasilia), Professor Ingetraut Dahlberg (Bad König), Professoressa Anna Dell'Agata (Pineto), David Fabish (Toronto), Ing. Udo Jauernig (Leonberg), Rakesh Jethwa (Toronto), Dr Eric McLuhan (Toronto), Mgr. Franz Nahrada (Vienna), Nino Nien (Maastricht), Dr Alan Radley (Manchester); Jonathan Robin (Paris), Professor Alfredo Ronchi (Milan), Dr Sabine Solf (Maastricht), Professoressa Rossana Valenti (Naples); Professor M P Satija (Amritsar), whose generous invitation prompted this paper and Dr Marie Luise Zarnitz (Tübingen).

The practical dimensions of this work would not have been possible without the support of Russian students led by Vasily and Alexander Churanov, and a team including Andrew Kotov, Maksim Kuptsov, and Alexander Sapozhkov. I am extremely grateful to them for their support in developing a new vision.

Appendix 1. Basic Relations as described by Dahlberg (Unpublished Manuscript, 2008)

Note that the field of computer science typically focuses mainly on content-related relations 1 and 2, that is, 2 of the 12 basic relations.

Formal relations

1 Identity
2 Inclusion
3 Intersection
4 Disjunction

Categorial relations

		Abstract entities
1	Entities	Concrete entities
		Principles
		Quantity
2	Properties	Quality
		Relation (in the sense of comparison)
		Operation (active)
3	Activities	Process (procedure)
		State (passive, zero-activity)
		Time
4	Dimensions	Space
		Position

[66] ITU Report: *The Internet of Things*, Geneva, 2005: http://www.itu.int/osg/spu/publications/internetofthings/. Cf the Conference. *Internet of Things. Internet of the Future*, Nice, October 2008: <http://www.internet2008.fr/spip.php?article9>

Content related relations
1 Generic relation (is a)
2 Partitive relation (has a)
3 Opposition/complementary relation
4 Functional relation

Appendix 2. Ten Basic Choices

Choice 1 Knowledge
Choice 2 Quantity
Choice 3 Quality
Choice 4 Activities
Choice 5 Space
Choice 6 Time
Choice 7 Questions
Choice 8 Media
Choice 9 Education
Choice 10 Access

Appendix 3. Choices 3 entails 12 Layers of Knowledge (K Choices)[67]

1 Terms	Classifications, Thesauri
2 Definitions	Dictionaries
3 Explanations	Encyclopaedias
4 Titles	Bibliographies, Catalogues
5 Partial Contents	Abstracts, Reviews

Primary Literature in Digital Library
6 Full Contents

Secondary Literature in Digital Library

7 Texts, Objects	Analyses, Interpretation, Criticism
8 Comparisons	Comparative Studies, Parallels
9 Interventions in Extant Object	Conservation
10 Studies of Non-Extant Object	Reconstructions

Future Primary/Secondary Literature (Virtual Agora)
11 Collaborative Discussions of Contents, Texts, Comparisons, Interventions, Studies
12 E-Preprints of Primary and Secondary Literature in Collaborative Contexts

It is useful to compare the above model with developments in Wiki, which began as an online electronic encyclopaedia (level 3). The past three years have seen on at least six levels:

1 Terms	Wikispecies, MetaWiki
2 Definition	Wiktionary
3 Explanations	Wikipedia
4 Titles	Wikipedia: Selected Works, Works cited, References Further reading
5 Partial contents	Wikiquote
6 Full contents	Wikibooks (including Wikijunior), Wikisource, Wikimedia Commons, Wikinews

Appendix 4. Basic searches entail Ten Goals Choices (G Choices):

1 Everyday
1a Emergency
2 Business
3 Culture
4 Learning
5 Environment
6 Government
7 Health
8 Legal
9 Leisure, tourism
10 Religion

[67] This first part of this list is adapted from figure 2 in the 2005 publication listed in note 45 above.

Appendix 5. Advanced work entails Ten Strategy Choices (S Choices)

1 Internal
2 Internet
3 Knowledge package
4 Proprietary database
5 Library subjects
6 Library classification
7 Library networks
8 Specialist collections
9 Relations
10 Emergence

Appendix 6. 10 Ten Layers of Terms (T Choices) in Research Mode

1 Universal terms
2 General terms
3 Personal terms
4 Field terms (in a database)
5 Subject terms
6 Classification terms
7 Classifications terms (in multiple systems)
8 Ontologies terms
9 Relations terms
10 Emergent term

They also reflect

Appendix 7: Ten kinds of Sources: (SO Choices)

1 Sample sources
2 Organised at random
3 Organised informally
4 Organised in proprietary form
5 Organised officially
6 Organised in 1 system
7 Organised in several systems
8 Organised for specialised fields
9 Organised with systematic relations
10 Organised by emergent patterns

FORM IV

(as per Rule 8)

Statement regarding ownership and other particulars about:
***WDL* (World Digital Libraries: An International Journal)**

1.	Place of Publication	: New Delhi
2.	Periodicity of its Publication	: Half-yearly
3.	Printer's Name	: Dr R K Pachauri
	Whether Citizen of India	: Yes
	Address:	The Energy and Resources Institute Darbari Seth Block IHC Complex, Lodhi Road New Delhi – 110 003
4.	Publisher's Name	: Dr R K Pachauri
	Whether Citizen of India	: Yes
	Address:	The Energy and Resources Institute Darbari Seth Block IHC Complex, Lodhi Road New Delhi – 110 003
5.	Editor's Name	: Debal Chandra Kar
	Whether Citizen of India	: Yes
	Address:	The Energy and Resources Institute Darbari Seth Block IHC Complex, Lodhi Road New Delhi – 110 003
6.	Name and address of individuals who own the newspaper and partners or shareholders holding more than one per cent of the total capital	: The Energy and Resources Institute Darbari Seth Block IHC Complex, Lodhi Road New Delhi – 110 003

I, R K Pachauri, hereby declare that the particulars given above are true to the best of my knowledge and belief.

Signature of Publisher
(Sd/-)
Dr R K Pachauri

Date: 1 July 2009

Who guards the guards: meeting the challenges of digital preservation

Heather Brown
Assistant Director, Paper, Artlab Australia and Project Officer, State Library of South Australia for University of S A Business Information Course

World Digital Libraries 2(1): 51–68

Abstract

This paper discusses the importance of preserving digital knowledge over a period of time, the challenges, priorities and the various principles and strategies for long term preservation. It draws on a risk management framework, using the analogy of 'who guards the guards' to explore the implications of dangers, the roles of those who safeguard digital content, and various strategies – technological, resourcing and organizational – for mitigating risks over a period of time. Three key strategies potentially have the capacity to profoundly influence the survival of digital knowledge into the future: collaboration and alliances, risk management and quality assurance, and education and training.

E-mail Brown.Heather@dpc.sa.gov.au

Digital preservation – the 'why'

Preservation links our past with the future.

As the world around us is changing rapidly, one of the few things that will last is the knowledge and cultural heritage that we preserve for future generations, as expressed in the following poem.

Knowledge

Stories can be read
in the mind's eye
without a single step of motion.

Lives live their own stories
within the creation of empires
that disappear on the wings
of distant memories.

What colour once ruled the skies
to be now seen in the horizon
of the setting sun?

Thunder rolls
the troops of conquest
until they become
the dust of another age.

Another face, another stage,
yet the story stays the same;
for what remains is
the breath that gives
the voice to all stories –
knowledge.

(©K S Brown 2005)

A significant part of the world's knowledge and heritage is now in digital form. As UNESCO has recognized in its Charter on the Preservation of Digital Heritage, this needs to be preserved as part of the cultural memory of the whole world (UNESCO 2004).

Allied with the cultural reasons are pragmatic, economic imperatives for preserving digital information. As Lynne Brindley, CEO of the British Library, comments: 'there are sound economic reasons for preserving the 'digital assets' that are resulting from large investments in digitising projects' (Brindley 2000).

Archivists and records managers will also attest that digital preservation is linked to accountability – particularly of governments and organizations. This view is supported by the NSF-DELOS Working Group on Digital Archiving and Preservation: '…we expect this [digital] content will remain accessible to allow us to validate claims, trace what we have done, or pass a record to future generations.'

So for all these reasons, the preservation of digital content is vitally important to our globally-linked, online, 24/7 societies.

The author uses the analogy of the 'troops of conquest' from the poem to symbolise the dangers faced by digital content over a period of time, and will explore the potential implications for the role of 'guarding' this content for the future.

Within this framework, the author uses the UNESCO definition of digital preservation as: 'the processes of maintaining accessibility of digital objects over time' (UNESCO 2003).

Digital content – the what

There are two streams to digital content.

On the one hand there is content that is 'born digital' such as websites and e-mails. On the other hand there is content that is 'turned digital', or copied ('reformatted') to digital, from paper-based, film-based or other analogue media.

In both of these streams, the digital formats and carriers are many and varied, and rapidly increasing in complexity and quantity. The 'bits' and 'bytes' can be delivered via a bewildering variety of magnetic tapes and disks, sound and image files, e-journals and interactive digital maps, databases, web sites and dynamic, mixed-media productions.

Both the sheer scale and variety of digital forms create key challenges for would-be guardians of digital repositories. Robin Dale

has likened the situation to a veritable 'wild west' of different digital repositories with variant file formats and platforms for maintaining and storing digital data (Dale 2006).

The dangers

These challenges in preserving digital content have been evocatively described as 'the digital dark age' (Deegan and Tanner 2002) and 'the dark archival underbelly' of 'the wonderful world of digitised information and online everything.' (US-Inter PARES Project, 2002 Quoting The Boston Globe 10 February 2000, cited in Harvey 2005).

In a nutshell, the 'troops of conquest' have taken on beguiling forms that infiltrate via changes in software and hardware. Furthermore, the old guardianship approach of 'benign neglect' that was previously applied to traditional library materials will not work with digital content. A new approach requiring frequent and active intervention is needed.

An infamous case study is that of the BBC Domesday project which produced a multi-media version of the Domesday Book on videodisc, to mark the 900th version of the original Domesday Book. Unlike the original book that has lasted over 900 years, the videodisc became inaccessible after just a few years in the late 1980s as the BBC microcomputer on which it was developed became obsolete. The data was later rescued using intensive emulation techniques (Abbott 2003).

The lack of reliable data about the costs of preserving digital information provides another set of potential danger.

Jonas Palm, Director, Head of Preservation at the National Archives of Sweden, aptly describes this in an article entitled 'The digital black hole': 'In the excitement about the solutions digitalization offers, the right questions about costs are often not asked, especially about the long term costs for keeping the digital files alive. This enthusiastic attitude is risky, for the conversion process to create the digital files may well be quite expensive to start with, and these investments may turn out to be wasted if planning for the future is ignored and no structural funding for maintenance is secured.… The more information is converted, the more costs for accessing it will go up. The digital black hole has got its firm grip on the project.'

While these are difficult to predict, overall there is a general agreement that they are significant. Jonas Palm clearly highlights that the real costs are in the ongoing management – the costs of 'guarding' the digital content for the future.

The issue of authenticity is another insidious, lurking danger – what attributes of digital materials do we preserve and what level of change, or loss is acceptable? This is particularly a dilemma in the case of 'born digital' material. For example, many websites are dynamic and interactive with 'search and retrieval aspects intrinsically bound up with their content'. However, we need to preserve the functionality of the search and retrieval components, as well as the data that the functionality interacts with to generate the required output (Smith 2003 quoted in Harvey 2005).

Confidence in the authenticity of digital content over a period of time is critical, owing to the ease with which alterations can be made, and the changes caused by processes such as migrating information from one system or carrier to another (UNESCO 2003, p. 113).

It is said that 'trust is like a crystal ball – once it is broken, you cannot fix it'. As India's IGNCA (Indira Gandhi National Centre for the Arts) and NMM (National Mission for Manuscripts) would attest, trust in authenticity is crucial in a country like India, with so much of its cultural heritage contained

within its ancient manuscripts. (IGNCA 2009, NMM 2009).

Allied to trust is the danger that some organizations responsible for digital libraries and repositories will inevitably cease to exist over a period of time, due to funding, political or social changes. Will their digital content be lost? Will it be technically possible to safeguard the content by transferring it to another organization with differing methods of storing and maintaining data? Similar issues are posed by threats of major disasters, terrorism, attacks by hackers or viruses, and civil unrest – particularly as they all affect power supplies and the integrity of buildings.

Another danger is that of ignorance – lack of awareness about the need for pro-active thinking and taking early and repeated actions to preserve digital materials. IT vendors demonstrated this lack of understanding during the BBC Domesday project:

> [one vendor] 'offered us a special polymer that they guaranteed would preserve a CD-ROM [disc] for 100 years. They were unable to answer how we would preserve a CD-ROM player for that length of time…' (Abbott 2003, p. 10)

Lack of established legal frameworks for copying digital materials for preservation purposes are another kind of threat.

'Copyright legislation impacts on the ability of organisations to undertake digital preservation activities. This can affect most areas of digital preservation, including copying and reformatting, preserving and providing long term access to stored material in archives, repositories and libraries. The problem is compounded by the lack of uniformity in copyright laws over various jurisdictions' (PADI 2009). As an example, Australia's Copyright Amendment (Digital Agenda) Act 2000 allows digital copying of materials held in libraries for preservation purposes, but in other countries, commercial imperatives to protect intellectual property rights may override preservation needs. Another layer of complexity is that of copyright in traditional knowledge systems that are embedded in digital objects (Gaur 2009).

A further threat is the imperative of organizations such as libraries and archives to provide access to knowledge and cultural heritage. Access is critical – without it collections are useless. Yet, particularly with 'turned digital' materials, this causes problems when digitization for access is also equated with long-term preservation.

When funding is directed simply to digitization for short-term access without consideration of the bigger preservation picture – including preserving the originals – we are in danger of losing our entire heritage; digital and non-digital.

Paolo Usai, the Director of Australia's Screensound, summed up this danger at an international Web Archiving Conference: 'If there is too much emphasis on access instead of preservation we lose balance and our cultural assets' (National Library of Australia 2004).

What should be kept?

On the other hand, destruction of some aspects of our digital culture is inevitable and even to be welcomed – due to the vast quantity and varying quality of what is created digitally on a daily basis. We simply cannot or even would want to preserve everything.

Ideally selecting for preservation is a pro-active process, involving ongoing interventions to make informed choices about what should travel across time for the benefit of future generations – a survival of the 'culturally-significant'.

According to the UNESCO Guidelines – on one level, the selection of digital heritage is conceptually the same as selection of non-digital material. In the best of worlds, deciding what to preserve should be linked to collection

management priorities and/or archival appraisal criteria and records management policies (UNESCO 2003).

However, on another level, digital materials present some new selection challenges. These include the sheer volume of materials and varying quality, and the complications of new emerging genres, together with the need to act quickly. According to the UNESCO Guidelines: 'It may not be possible to wait for evidence of enduring value to emerge before making selection decisions.'

There is also the challenge of deciding how much to select. In web archiving for example, should the approach be selective as with the National Library of Australia's PANDORA model, or the whole of domain snapshot harvesting of the Internet Archive, or a combination?

It is inevitably a balancing act. While we cannot preserve the whole digital universe, the UNESCO Guidelines suggest that it is probably better to err on the side of collecting more material than less. In selecting for the future the UNESCO Guidelines provide some useful signposts, including the following.

- Decisions should be based primarily on the value of material in supporting the mission of the organization.
- This value must be weighed against the costs and difficulties of preservation and availability of resources.
- It is desirable to preserve at least a sample of all kinds of digital materials, including the clearly ephemeral.

Another helpful resource is the Decision Tree that outlines options and pathways for selecting what to preserve in the DPC Handbook. (DPC 2002)

Whatever the approach, the decision needs to be cautious and taken early for, as the UNESCO Guidelines state, a decision not to preserve is usually a final one for digital materials (UNESCO 2003, p. 75) .

So far, my focus has been on the why and what of preserving digital heritage, and on some of the key dangers. These issues set the scene for the next stage.

Who are the 'guards'?

Traditionally libraries and archives have seen the preservation of resources as a core responsibility, with the librarians, archivists and records managers taking on the role of 'safe-guarding.'

'Society has always created objects and records… and it has consciously preserved them in a permanent way… Cultural institutions are recognized custodians of this collective memory: archives, libraries and museums play a vital role in organizing, preserving and providing access to the cultural intellectual and historical resources of society. They have established formal preservation programs for traditional materials and they understand how to safeguard both the contextual circumstances and the authenticity and integrity of information and objects placed in their care'
(Smith 2002, pp. 133–134).

With digital information, the safeguarding role is expanded to new dimensions. Librarians, archivists, and records managers, the traditional guardians of heritage, will also need to work collaboratively at an international level and intensively with other, newer stakeholders in preserving digital content. (Hoorens 2008). According to UNESCO's Charter on the Preservation of Digital Heritage these newer stakeholders potentially include a diverse team of: 'hardware and software developers, creators, publishers, producers and distributors of digital materials as well as other private sector partners' (UNESCO 2004 Article 10).

While considerable debate has focused on the roles these various stakeholders might take, ultimately it seems probable that digital preservation will be led by those stakeholders

with an interest in or mandated responsibility for preserving knowledge for the future.

So what does this mean for the roles of these new guards? And what does this mean in terms of monitoring their roles, or 'guarding the guards'?

To explore these issues, I now want to turn to the framework of risk management, drawing on the work of Jonathan Ashley-Smith (Ashley-Smith 1999).

Risk management: a framework for 'guarding the guards'

An analogy of risk

The well-known story of the old Indian city of Daulatabad is a powerful reminder about risks, and it is a story that has lessons for all of us who are in the business of trying to preserve knowledge over a period of time. In a nutshell the ruler wanted to build a fort that would safeguard his people – a most trustworthy, impregnable fort, with a huge moat and spikes on the gates to ward off elephant charges. The impregnable fort was invaded – simply – by bribing the guard at the gate.

From a risk-management perspective, the potential dangers outlined earlier are viewed as 'risk factors' and they are then assessed in terms of significance, likelihood and consequences.

Working within this framework we can now turn to some principles and strategies that may help to mitigate the risks and to 'guard the guards.'

Principles and strategies

It is clear that there is no single magic weapon or shield to help solve the complex, challenging problems of digital preservation. It is likely to be a combination of several or all of these strategies integrated into the day to day activities of key command centres, staffed by a wide range of digital stakeholders, that will succeed in achieving some level of risk mitigation.

Standards, guidelines, and preservation metadata

Foremost among these strategies are guidelines and standards that play a key role in reducing the risk of losing digital content. They promote certain common requirements for quality and also make it easier for collaboration and interoperability among organizations.

Good progress has been made in a number of areas. These include identifying and promoting particular formats and features that are likely to make preservation easier – for example the TIFF and JPEG 2000 image file formats. (Buckley 2008). Also useful are the ISO standards for preservation of digital materials (Adelstein 2003) and the UNESCO Guidelines (UNESCO 2003).

OAIS model

Potentially one of the most important frameworks developed to date is the OAIS (Open Archive Information System) reference model. The model is an ISO standard and provides a common framework for describing and comparing architectures and operations of digital archives and is used as the basis for a wide range of digital preservation systems internationally (CCSDS 2002).

Preservation metadata

The concept of preservation metadata underpins this framework. Preservation metadata describes and records information needed to manage the preservation of digital resources; information such as the date of capture, the capture device and change history. The PREMIS metadata model is entwined throughout the OAIS framework and is designed to be applied flexibly across different types of systems (Guenther 2006, Woodyard-Robinson 2006).

This framework has now been developed to enable digital objects to be exchanged in a meaningful way between different platforms

(Raftos 2006). This is a vitally important risk mitigation strategy: for the first time it is now technically possible to safeguard digital content by transferring it to another organization with differing platforms.

Other features – unique identifiers, encapsulation, and others

Linked with the OAIS framework and preservation metadata are other features that help ensure ongoing access to authentic and trusted digital objects.

Among these are the measures for using unique identifiers. These are needed because the current system of identifiers (URLs) is based on file location, and so a change in location means there is a need to change in identifier. Like the gateways themselves, new systems of persistent identifiers are continually evolving; the National Library of Australia has recently adopted a new model of persistent identifiers for its newspaper digitalization project (NLA 2007).

Encapsulation techniques also assist ongoing access, and are linked with emulation. These techniques relate to the packaging together of the digital item, its preservation metadata and other associated data – possibly even the associated software required for access. The packaging process lessens the likelihood that any key component needed to decode a digital object will be lost.

Other measures that can help ensure authenticity include digital watermarking and signatures, encryption, digital time stamping, audit trails, and controlled custody (NSF-DELOS Working Group on Digital Archiving and Preservation, 2003 p. 6).

However, as Cullen points out, while, such strategies improve over a period of time, they may not ensure unconditional authenticity, and he poses the hard question: how confident can one be when an object whose authentication is crucial depends on electricity for its existence (Cullen 2000). Further research and development is needed in this area of vulnerability.

Auditing and Trusted Digital Repositories

Auditing is a powerful tool for checking the level of risk and 'guarding the guards'. A continuous auditing cycle makes sure systems are working to quality assurance levels and links back to their role in maintaining the identity, integrity, and quality of the digital copy as a trusted source of the cultural record.

A landmark for auditing is TRAC (Trustworthy Repositories Audit and Certification: Criteria and Checklist) produced by OCLC, the Center for Research Libraries and the National Archives and Records Administration. Revised in 2007, this is the first guide for objectively determining whether a digital repository can be a long-term trusted location for digital content.

Not surprisingly, the TRAC Checklist draws on a risk management framework and the OAIS model to identify indicators of trustworthiness and reliability for digital repositories to manage their digital resources to their 'designated communities' now and into the future.

A range of similar auditing tools is emerging, including DRAMBORA (The Digital Repository Audit Method Based on Risk Assessment). DRAMBORA is a toolkit for providing repository administrators with a self-check framework. Similarly, NESTOR, the agency assigned to the task of auditing and certifying digital archives in Germany, has developed its own auditing tools.

As an example of the auditing process in action, during 2009, the Center for Research Libraries audited two repositories with the goal of their certification as trustworthy digital repositories (CRL 2009).

Overall, such a trusted digital repository is 'a complex interrelated system'. (NESTOR 2006).

Key features include: 'the organization running the repository: its governance, organizational structure and staffing; policies and procedures; financial fitness and sustainability; the contracts, licenses, and liabilities under which it must operate; and trusted inheritors of data, as applicable. Additionally, the digital object management practices, technological infrastructure, and data security in place must be reasonable and adequate to fulfil the mission and commitments of the repository' (TRAC 2007 Introduction).

A trusted digital repository will also recognise threats to and risks within its systems and will undergo constant monitoring, planning, and maintenance to carry out its mission of digital preservation.

What is clear is that these complex, interrelated systems will require frequent, ongoing auditing, and that continuing research and development is required to refine the auditing tools for risk analysis.

Refreshing, migrating, emulation – technological strategies

In combination, three key technological strategies are a powerful means of mitigating risk to digital content if they are used within a systematic framework such as the OAIS model.

They comprise:

- refreshing – periodically moving a file from one physical storage medium to another as an ongoing process
- migration – moving files from one file encoding format to another
- emulation – recreating the application environments on which the original files can run.

To make these strategies more effective, new systems architectures and tools such as PANIC and AONS are being developed to automatically detect obsolescence and notify that actions are needed (Hunter 2006).

The European consortium PLANETS is providing a 'test bed', a controlled environment where such strategies and tools can be tested and evaluated (PLANETS 2009).

Microfilm – a prudent option for text-based content

In the case of 'turned digital' content that is derived from text-based materials, another prudent risk management strategy is the so-called hybrid option. A microfilm copy provides for long term preservation, together with a digital copy providing the flexible, multidimensional access.

Microfilm produced and stored to well-established, rigorous international quality assurance standards has a life-expectancy of 500 years.

This hybrid approach meshes with Robin Dale's comments that, as different repositories keep their data in different ways, it is important to use a variety of preservation options within and between organisations (Dale 2006).

The hybrid approach has also been endorsed by IFLA where microfilm has become an integral step in newspaper digitisation programs (IFLA 2002).

From a financial perspective, combining microfilming and digitizing might be seen as costly and unnecessary duplication. The reality is that it provides a flexible, economic coalition. The microfilm copy provides a platform that is low risk and from which it is frequently more cost effective to digitize than from the originals (Brown and Fenton 2005).

Jonas Palm also supports the microfilming option. In The Digital Black Hole he comments that the hybrid solution of COM (computer output microfilm) along with the digital copy may, in financial terms, be a more prudent strategy.

The Swedish National Archives [Riksarkivet] is currently studying whether it is feasible to use COM in an effort to improve the strategy of microfilming, which has a long record for

securing information on materials in bad condition (Palm 2006).

It is also a strategy adopted by the National Archives of Singapore which captures its electronic records and immediately writes them to microfilm as the long term option (National Archives of Singapore 2009).

Legislation

To help get hold of digital materials in the first place a number of countries have, or are currently considering, legal deposit legislation that covers digital materials. New Zealand, for example has passed legislation that 'provides for the deposit for physical format documents and for the copying of internet documents' and Australia is moving in the same direction (Verheul 2006).

Similarly, legislation to allow copying of digital materials for preservation purposes is a key strategy to ensure that vital processes such as refreshing, migrating, emulation are not curtailed by legal requirements. As the PADI website reports: '…Solutions include introducing licencing for preservation and lobbying governments to change legislation so that there is a balance between the rights of the copyright holders to protect their interests and exemptions for institutions for legitimate long term preservation purposes' (PADI 2009).

Resourcing

As the Cornell Digital Preservation Tutorial highlights, adequate resourcing is just as critical as the technological strategies in ensuring preservation of digital content over a period of time. Building a long term resourcing infrastructure – finding ongoing funding to sustain the guards and systems – is particularly challenging (Shenton 2009). For example, while it is likely that unit data storage costs will decline over a period of time, the overall volume of data to be stored will continue to grow exponentially in sheer numbers, and as digital objects incorporate more features.

Jonas Palm evocatively describes this dilemma:

> 'In the excitement about the solutions digitization offers, the right questions about costs are often not asked, especially about the long term costs for keeping the digital files alive. This enthusiastic attitude is risky, for the conversion process to create the digital files may well be quite expensive to start with, and these investments may turn out to be wasted if planning for the future is ignored and no structural funding for maintenance is secured.
>
> Without such long-term planning, digitization projects can come to behave like black holes in the sky' (Palm 2006).

Colin Webb describes how difficult it is to argue the case for ongoing funding to administrators when the long-term costs are unknown. 'To say "we only know it will cost a lot" is an unsatisfactory answer, even if it is the truth' (Webb 2004 p. 45).

However, doing nothing – or paralysis is not an option. Very useful steps for moving forward are outlined in the Cornell Digital Preservation Tutorial including identifying start and ongoing costs. It is often easier to obtain seed money for start up costs. Ongoing costs need to be found from a variety of strategies. At the National Library of Australia, it has been possible to find some funding for ongoing costs by reallocating some priorities and drawing on skills and commitment already in the organization. In Webb's view: 'this has tended to encourage senior managers to pay close attention to digital preservation programmes, and help embed such programmes in the core business of the library.' (Webb 2004 p. 45)

Organizational infrastructure

Organizational infrastructure is equally critical to safeguarding digital heritage. Overall, it involves organizational commitment as reflected through the policies and procedures and the plans for digital preservation. This means writing and supporting, implementing, reviewing and maintaining policies, plans and procedures and getting 'buy in' right from the top. Without this high level commitment, the attempt to guard digital content over time will lack direction and ultimately fail.

Policy direction aligns the strategies for preserving digital information and applies resources where they are most needed. Examples of high level policies and strategies are the National Library of Australia's Digital Preservation Policy and the British Library's Digital Preservation Strategy. India is similarly developing its policies and strategies (IGNCA 2009). Inevitably these will link with selection criteria – the decisions about what to preserve, as discussed earlier. They also deal with roles and responsibilities and managing intellectual property. There are close links with technological infrastructure with requirements for managing quality assurance.

Collaboration

As preservation challenges are the same for everyone in the digital universe, collaboration is a way of sharing expertise in the construction and ongoing maintenance of the gateways. Inevitably these become part of the organizational infrastructure. More formal agreements and collaborations are an important way of sharing development costs, harnessing and focusing effort, and attracting resourcing and support for programs (UNESCO 2003 pp. 64–65).

Examples at the international level include alliances and services such as ICADS (a joint alliance of IFLA and the Conference of Directors of National Libraries), UNESCO, OCLC, PADI, PLANETS, The Internet Archive and The International Internet Preservation Consortium. Verheul highlights many such examples at regional and national levels in the 2006 IFLA survey Networking for Digital Preservation (Verheul 2006).

However, while collaboration and alliances are beneficial, they also cost and need ongoing commitment. As Webb comments: 'the most effective collaboration programmes seem to have been based on areas of real common interest, realistic expectations, clear understanding about who is responsible for what, and the allocation of sufficient resources to pay attention to the collaborative relationship itself' (Webb 2004).

Education and training

Training in quality assurance and in preservation strategies – training the guards and other stakeholders – is a fundamental risk management strategy. It impacts directly on monitoring the auditors or 'guarding the guards'.

Whole new skill sets are needed by digital librarians and other stakeholders to develop and sustain digital preservation. Not only do the new guards need an awareness of the challenges and technological knowledge, they also need more generic skills that particularly impact on the organizational and resourcing infrastructure, skills such as project management and communication skills and the ability to adapt to change. They also need vision and critical thinking skills together with the knowledge and ability to deal with technological challenges; to proactively select, intervene, and to collaborate across boundaries in order to build preservation and sustain gateways for the long term.

The challenge is for the universities and educators of librarians to provide the relevant foundation of knowledge and skills, and for digital library staff to continue to develop and

adapt their knowledge and skills as part of their life-long learning. (Rasmussen and Youngok 2006).

Education and training can take the form of formal qualifications, workshops, short courses, conferences and self-learning, and all have their place in the continuous learning spectrum.

An example of a well established formal training programme is HATII (the Humanities Advanced Technology and Information Institute) at the University of Glasgow that offers a Masters in Information Management and Preservation – with a focus on digital preservation (HATII 2009).

As an example of formal training within Australia, I am currently involved in developing part of the new BIM (Business Information Management) Course at the University of South Australia that is taking an innovative approach. Recognising that the information world is one of constant change and blurring boundaries of responsibility, the new program integrates the areas of archives, records management, knowledge management, librarianship and business information. The aim of the course is to develop awareness of a wide range of stakeholders to key issues in information management such as digital preservation, which are considered within a risk management framework (University of South Australia2009). More formal programmes like these are needed across the region.

As an example of a workshop model within the region, the National Archives of Singapore continues to provide leading-edge training workshops to regional colleagues about digitizing and microfilming quality issues to practitioners and project managers, focusing on the risks and benefits of different approaches.

As examples within India, DELNET plays a leading role in providing training and professional development to library staff – including digital preservation (DELNET 2009). Likewise, the ICDL Conference on digital libraries is increasingly addressing issues of digital sustainability and preservation (ICDL 2006).

To complement these, a wide range of training courses are now available that not only raise awareness of quality issues but also highlight the risks in preserving digital content.

Some well-known online versions include:

- Digital Preservation Tutorial and Survey of Institutional Readiness from Cornell University.
- Digital Imaging Tutorial from LYRASIS.

A valuable support for life-long learning is the National Library of Australia's PADI (Preserving Access to Digital Information) site that provides a valuable educational framework and signpost relating to digital preservation issues (PADI 2009).

In summary, more capacity building in the area of digital preservation is needed across the region. Training provides a pathway of skills and experience critical to monitoring the guards. Training can also equip future guards with the vision, critical thinking skills, and ability to deal with future 'troops of conquest'; to proactively select, intervene, and collaborate across boundaries in order to safeguard digital heritage in the long term.

Conclusion

So much has been achieved in developing technological, organizational and resourcing strategies that mitigate the risks to digital content over a period of time. Applied in combination and with variation in approaches, they can potentially be a powerful defence against the dangers of digital destruction.

Notwithstanding these achievements, right now the risks in digital preservation still remains high, and the likelihood and consequences of loss are still dire – because these strategies are still in the early stages of

development and testing (Hoorens and Rothernberg 2008; Shenton 2009).

Or, to use Veheul's words: 'Digital preservation is still under construction' (Verheul 2006, p. 69).

Nonetheless, as we continue to build our globally interlinked digital repositories, there are three interlinked and absolutely pivotal options, which is believed to have the capacity to profoundly influence the survival of digital knowledge into the future. These are:

- Collaboration and alliances
- Risk management, quality assurance auditing and certification
- Education and training

As Colin Webb from the National Library of Australia insightfully comments: 'An effective preservation role in the library of the future will require flexibility, willingness to change, proactively seeking a useful role that draws on the expertise and perspectives we already have, while developing whatever new expertise and perspectives will be needed. Willingness to listen, consult, learn, and to form alliances and partnerships will all be important' (Webb 2002).

'Another face, another stage' and the story continues through time. The above three options, in combination, and using a variety of strategies, will effectively provide a framework for preservation of digital knowledge for the future.

Applying risk management systems, they will equip stakeholders with the vision and skills to rigorously and objectively keep on re-assessing that quintessential question: Who Guards the Guards?

References

Abbott D. 2003
Overcoming the dangers of technological obsolescence: rescuing the BBC Domesday project
DigiCULT.Info: a newsletter on digital culture, **4**, August, pp. 7–10

Adelstein P Z. 2003
Preservation of electronic records: status of ISO standards'
In *Symposium* 2003: *Preservation of electronic records*: *New Knowledge and Decision Making*
Ottawa, 15–18 December 2003: preprints

Ashley–Smith J. 1999
Risk assessment for object conservation
Oxford, Butterworth-Heinemann

Brindley L J. 2000
Keynote address to the Preservation 2000 conference.
New Review of Academic Librarianship **6:** 125–137

Brown H and Fenton A. 2005
Time's pivotal point – preserving the past for the future now
In *International Newspaper Librarianship for the 21st Century,* edited by H Walravens
IFLA Newspapers section. International Newspapers Conference: Asia and the Pacific: Newspapers in the 21st Century 22–24 February. National Library of Australia. Berlin, Saur, pp. 63–72. Details available at <http://www.reference-global.com/doi/abs/10.1515/9783598440205.2.63>, last accessed on 12 April 2009

Brown K S. 2005
Knowledge [unpublished poem]
K S Brown 13 September 2005

Buckley R. 2008
JPEG 2000 – a practical digital preservation standard? DPC Technology Watch Series report 08-01, Digital Preservation Coalition. Details available at <http://www.dpconline.org/docs/reports/dpctw08-01.pdf>, last accessed on 12 April 2009

CCSDS (Consultative Committee for Space Data Systems). 2002
Reference Model for an Open Archival Information System, (ISO Standard 14721).
Details available at <www.ccsds.org/publications/archive/650x0b1.pdf> last accessed on 2 April 2009

CHIN (Canadian Heritage Information Network). 2009
Capture your collections - planning and implementing digitization projects [online tutorial] CHIN
Details available at <http://www.chin.gc.ca/English/Digital_Content/Capture_Collections/index.html>, last accessed on 2 April 2009

Cornell University Library. 2009
Digital Preservation Tutorial, Details available at <www.library.cornell.edu/iris/tutorial/dpm/>,
Digital Timeline, Details available at <http://www.library.cornell.edu/iris/tutorial/dpm/timeline/viewall.html>,
last accessed on 2 April 2009

CRL (Center for Research Libraries). 2009
CRL's Certification and Assessment of Portico and Hathi Trust
Details available at <http://www.crl.edu/content.asp?l1=13&l2=58&l3=181> last accessed on 12 April 2009

Cullen C. 2000
Authentication of digital objects: lessons from an historian's research
In CLIR (Council on Library and Information Resources)
Authenticity in a Digital Environment, Washington, CLIR pp. 1–7

Dale R. 2006
Certifying Digital Repositories: Development of the Audit Checklist [and summing up comments] in APSR [Australian Partnerships for Sustainable Repositories]. *Long term repositories: taking the shock out of the future* 31 August–1 September 2006. Details available at <www.apsr.edu.au/presentations/longterm.html>, last accessed on 12 April 2009
Canberra: APSR

Deegan M and Tanner S. 2002
In *The Digital Dark Ages*, Update May 2002
Details available at <www.cilip.org.uk/publications/updatemagazine/archive/archive2002/may/update0205b.htm> last accessed on 2 April 2009

DELNET
Details available at <http://www.delnet.nic.in> last accessed on 2 April 2009

DPC (Digital Preservation Coalition). 2002
Preservation Management of Digital Materials – the Handbook by N Beagrie and M Jones Digital Preservation Coalition, UK (and decision tree) Details available at <www.dpconline.org/graphics/handbook/.htll>

Gaur R. 2009
Barriers to access Indian cultural heritage: copyright, permission rights and ownership issues vis-à-vis IGNCA collections
In Proceedings Indo-US Workshop on international trends in digital preservation, March 24–25 2009, Human Centred Design and Computing, CDAC, Pune, India pp. 20–30. Details available at <http://www.ndpp.in/indo_us.html>, last accessed 15 July 2009.

Guenther R. 2006
The promise of PREMIS: background, scope and purpose of the 'Data Dictionary for Preservation Metadata' in APSR [Australian Partnerships for Sustainable Repositories]
Long term repositories: taking the shock out of the future. 31 August–1 September 2006, Canberra, APSR
Details available at <http://www.apsr.edu.au/presentations/longterm.html>, last accessed on 12 April 2009

Harvey R. 2005
Preserving Digital Materials
Munchen: K G Saur

HATII (the Humanities Advanced Technology and Information Institute) at the University of Glasgow
Details available at <http://www.hatii.arts.gla.ac.uk>, last accessed on 2 April 2009

Hoorens S and Rothernberg J. 2008
Digital preservation: the uncertain future of saving the past RAND Project resource, RAND Europe
Details available at <http://www.rand.org/pubs/research_briefs/RB9331/> last accessed on 12 April 2009

Hunter J. 2006
Putting the shock back into the future
In APSR [Australian Partnerships for Sustainable Repositories]
Long term repositories: taking the shock out of the future 31 August–1 September 2006, Canberra, APSR
Details available at <http://www.apsr.edu.au/presentations/longterm.html., last accessed on 12 April 2009

ICDL (International Conference on Digital Libraries). 2006
Details available at <http://www.teriin.org/events/icdl/> last accessed 12 April 2009
New Delhi: The Energy and Resources Institute

IFLA Newspapers section
- 1996, *Guidelines for Newspaper Preservation Microfilming*
- 2002, Microfilming for Digitisation and Optical Character Recognition: Supplement to Guidelines

Details available at <http://www.ifla.org/VII/s39/broch/microfilming.htm>, last accessed on 12 April 2009

IGNCA (Indira Gandhi National Centre for the Arts). 2009
Seminar on Digital Preservation and Access to Indian Cultural Heritage with special reference to IGNCA Cultural Knowledge Resources 31 January–1 February 2009
Details available at <http://ignca.nic.in/kn_digital_program.htm>

National Archives of Singapore. 2009
Details available at <http://www.nhb.gov.sg> last accessed on 12 April 2009

National Library of Australia
- Web Archiving Conference November 2004
  Details available at <http://www.nla.gov.au/>, last accessed on 2 April 2009
- Australian Newspapers Digitisation Program. 2007 National Library of Australia Persistent Identification Model for Newspapers
  Details available at <http://www.nla.gov.au/ndp/project_details/documents/ANDP_Newspaper_PIs_v1.0p>
- Digital Preservation Policy 2008 (3rd ed)
  Details available at <http://www.nla.gov.au/policy/digpres.html> last accessed on 2 April 2009

NNM (National Mission for Manuscripts [India]). 2009
Details available at <http://namami.org/index.htm> last accessed on 2 April 2009

NESTOR. 2006
Details available at <http://www.langzeitarchivierung.de/> last accessed on 12 April 2009

Nestor Working Group on Trusted Repositories Certification. June 2006
Catalogue of Criteria for Trusted Digital Repositories. Version 1 (draft for public comment)
English translation December 2006. urn:nbn:de:0008-2006060703.
edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

NSF-DELOS Working Group on Digital Archiving and Preservation. 2003
Invest to save: report and recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation
National Science Foundation and the European Union

OCLC and Research Libraries Group (RLG). 2002
Trusted Digital Repositories: Attributes and Responsibilities
Details available at <http://www.oclc.org/au/en/digitalpreservation/resources/default.htm> last accessed on 2 April 2009

PADI (Preserving Access to Digital information). 2009
Details available at <http://www.nla.gov.au/padi/> last accessed on 2 April 2009

Palm J. 2006
The digital black hole
[Jonas Palm, Director, Head of Department of Preservation, Riksarvet/National Archives, Stockholm Sweden – article posted May 2006]
Details available at <http://www.tape-online.net/docs/Palm_Black-Hole.pdf> last accessed on 14 June 2006
The article is published in the framework of the TAPE project: Training for Audiovisual Preservation in Europe, <http://www.tape-online.net/>

Pandora
Details available at <http://www.nla.gov.au/pandora/> last accessed on 2 April 2009

Phillips M. 2002
Archiving the web: the National Collection of Australian Online Publications
Paper presented at the International Symposium on Web Archiving, National Diet Library, Tokyo, Japan 2002.
Details available at <http://www.nla.gov.au/nla/staff-paper/2002/phillips1.html> last accessed on 2 April 2009

PLANETS (Preservation and Long-term Access through Networked Services). 2009
Details available at <http://www.planets-project.eu/> last accessed on 2 April 2009

Rassmussen E and Choi Youngok. 2006
Digital librarians for digital libraries
In *ICDL International Conference on Digital Libraries*,
New Delhi: TERI. pp. 748–754

Shenton H. 2009
Digital versus print as a preservation format – Expert views from international comparator libraries
Details available at <http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digitalvprint/index.html>, last accessed on 12 April 2009

Smith A. 2003
***New-model scholarship – how will it survive*?**
Washington, DC: Council on Library and Information Resources

Smith A. 2004
The future of web resources: who decides what gets saved and how do they do it
Paper delivered at international conference Archiving Web Resources: issues for cultural heritage institutions
9–11 November 2004
Canberra: National Library of Australia

Smith B. 2002
Preserving tomorrow's memory: preserving digital content for future generations
Information Services & Use, **22**: pp. 133–139

SOLINET Digital Imaging Tutorial
Details available at <http://www.solinet.net/workshops/workshop_templ.cfm?doc_id=1085> last accessed on 13 April 2009

State Library of South Australia
Details available at <http://www.slsa.sa.gov.au> last accessed on 13 April 2009

UNESCO. 2003
Guidelines for the Preservation of Digital Heritage, prepared by the National Library of Australia
Paris: UNESCO.
Details available at <http://www.unesdoc.unesco.org/images/0013/001300/13007le.pdf> last accessed on 13 April 2009

UNESCO. 2004
Charter on the Preservation of Digital Heritage
Paris: UNESCO
Details available at <http://www.unesco.org.ci/en/file_download.php/4cc126a2692a22c7c7dcc5ef2e2878c7Charter_en.pdf>, last accessed on 13 April 2009

University of South Australia. 2009
Business and Information Management course
Details available at <http://www.unisa.edu.au/bim/> last accessed on 12 April 2009

US-Inter PARES Project. 2002
InterPARES Interpreted: a guide to findings on the preservation of authentic electronic record
Details available at <http://www.gseis.ucla.edu/us-interpares/pdf/InterPARESInterpreted.pdf>, last accessed on 12 April 2009

Verheul I. 2006
Networking for Digital Preservation: Current Practice in 15 National Libraries (IFLA Publications 119)
Munchen: K G Saur.

Webb C. 2004
The malleability of fire: preserving digital information
In *Managing preservation for Libraries and archives: current practice and future developments*, edited by Feather, J Ashgate, Aldershot. pp. 27–52

Woodyard-Robinson D. 2006
Common use cases for preservation metadata
In APSR [Australian Partnerships for Sustainable Repositories]
Long term repositories: taking the shock out of the future, 31 August–1 September 2006
Canberra: APSR
Details available at <http://www.apsr.edu.au/presentations/longterm.html>, last accessed on 12 April 2009

Bibliography

Adams D, *et. al.* 2004
Addressing a risk perspective when considering digitisation as a preservation reformatting method: a response from 'down under'
Microform and Imaging Review **33**(4): Autumn pp. 188–189

Arthur K, *et al.* 2004
Recognizing Digitization as a Preservation Reformatting Method
Prepared for the ARL Preservation of Research Library Materials Committee
Details available at <http://www.arl.org/preserv/digit_final.html> last accessed on 2 April 2009

Beagrie N. 2004
The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee
In *D-Lib magazine*, **10**(7/8) July. Details available at <http://www.dlib.org/dlib/july04/beagrie/07beagrie.html> last accessed on 2 April 2009

British Library. 2007
British Library Digital Preservation Strategy
Details available at <http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf>

De Lusenet, Yola 2007
Tending the garden or harvesting the fields: digital preservation and the UNESCO Charter on the Preservation of Digital heritage
In *Library Trends Summer* **56**(1). pp.164–182

DRAMBORA (The Digital Repository Audit Method Based on Risk Assessment)
The Digital Repository Audit Method Based on Risk Assessment 2009
Details available at <http://www.repositoryaudit.eu/> last accessed on 12 April 2009

Harvey R. 1993
Preservation in Australian and New Zealand Libraries: principles, strategies and practice for librarians, 2nd ed
Wagga Wagga, Centre for Information Studies, Charles Sturt University

Hedstrom M. 1998
Digital preservation: a time bomb for digital libraries
In *Computers and Humanities,* **31**: pp. 189–202

Hodge G and Frangakis E. 2004
Digital preservation and permanent access to scientific information: the state of the practice
A report sponsored by the International Council for Scientific and Technical Information (ICSTI) and CEDNI
Details available at <http://cendi.dtic.mil/publications/04-3dig_preserv.html>, last accessed on 13 April 2009

ICADS [a joint alliance of IFLA and the Conference of Directors of National Libraries]
Details available at <http://www.nla.gov.au/padi/topics/712.html> last accessed on 9 April 2009

The Internet Archive
Details available at <http://www.archive.org>, last accessed on 2 April 2009

International Internet Preservation Consortium
Details available at <http://www.netpreserv.org> last accessed on 2 April 2009.

JISC (Joint Information Systems Committee). 2009
Digital Preservation Training Program
Details available at <http://www.ulcc.ac.uk/dptp/> last accessed on 2 April 2009

Kenney A, *et al.* 2006
E-Journal Archiving Metes and Bounds: A Survey of the Landscape
CLIR pub 138, CLIR September, 2006
Details available at <http://www.clir.org/PUBS/reports/pub138/pub138.pdf> last accessed on 12 April 2009

LYRASIS (formerly SOLINET) Distance Education courses. Digital Imaging of Library and Archival Materials
http://www.lyrasis.org/Classes-and-Events/Tracks/Digitization.aspx last accessed on 12 July 2009

Information provision to knowledge creation: Danish digital libraries strategy

Jakob Heide Petersen, Head of Division, and
Jens Thorhauge, Director General, Danish Agency for Libraries and Media

World Digital Libraries 2(1): 69–82

Abstract

This paper describes Danish digital libraries strategy in knowledge creation. While doing so, authors describe the role and issues for digital libraries. They have analysed the Danish deliberations regarding the development of digital library services in order to provide inspiration outside Denmark. They have described in detail history and strategies for development of e-services in public libraries in Denmark. They have also descried how and why Denmark's Electronic Research Library was established, via a network of cooperating electronic research libraries and how they developed digital library services in the library's environment. They also describe changes in education and learning, consequences for library development, developing a new role for the library, digital library services, concrete activities, and service development, institutional repositories and open access and organizational aspects – institutional partnerships in developing digital library in Danish point of view.

Introduction – role of, and issues for digital libraries

We often hear that libraries (and librarians) select, organize, retrieve, and transmit information or knowledge. That is true. But those are the activities, not the mission, of the library. … the important question is: 'To what purpose?' We do not do those things by and for themselves. We do them in order to address an important and continuing need of the society we seek to serve. In short, we do it to support learning.

Lorcan Dempsey, Keynote presentation at European Conference on Digital Libraries 2004[1]

As indicated in the above quote the focus in recent library development has been on new ways to fulfil the purpose of the library. In 2003 OCLC (Online Computer Library Center) carried out a very comprehensive study of the information environment and the library's situation in general. The subsequent report from this Pattern Recognition Scan highlighted the need for libraries to focus more on the overall mission and less on the activities and workflows traditionally used to achieve the goals of the library. The report contained an explicit criticism of the libraries' focus on their specific collections and a recommendation to focus more on user needs. These recommendations have since been strengthened by a number of reports from OCLC for instance on the libraries' catalogues and their image among users. The theme of these reports is not so much a questioning of the library's role or mission as a need for new approaches to support learning and research.

This issue has been the centre of Danish library development in recent years and of the attempts to formulate national policies and strategies on the subject. The Danish strategies and policies have been formulated on the basis of surveys and analyses carried out in Denmark that confirm general developments in end users' information habits.

The natural starting point for a discussion of library tasks is the library's traditional role in information provision. It is this role that is being challenged in the digital environment but it remains an important task for all libraries.

We will argue, however, that information provision is not the only important role for the library in the transmission of information through the value chain from author to end user. The individual library should consider the new tasks in relation to e-publishing and e-learning as being as important as the more traditional role of information provision. The task of information provision remains central but should probably be organised in a new way – changing the role of the individual library. You may argue that we can identify a need for enabling users to benefit better from their information access and that libraries at universities and other academic institutions tend to build institutional repositories and in consequence should promote the research carried out at their institutions.

The following presentation of Danish deliberations regarding the development of digital library services hopefully provides inspiration outside Denmark.

Strategies for development of e-services in public libraries

In 1995 the first Danish public library (in Silkeborg) offered access to the Internet for its users and at the same time presented a homepage. The immediate success of these services inspired many libraries to follow, and at the same time a discussion on the

[1] Lorcan Dempsey, Libraries, Digital libraries and digital library research, Keynote presentation at European Conference on Digital Libraries 2004, University of Bath 12–17 September 2004. Details available at <http://www.oclc.org/research/presentations/dempsey/ecdl2004.ppt>last accessed on 8 July 2005.

possibilities of the web-technology gained a very broad basis in the Danish libraries.

In 1996 the Danish National Library Authority (today DALM (Danish Agency for Libraries and Media) published guidelines and introduced some basic standards. A national programme was launched offering economic support to libraries in small municipalities to establish internet access and homepages.

The following year the discussions on extending e-services led to the first introduction of a Danish e-ask-a-librarian-service. The service was based in a few co-operating libraries. Today the service is based on a co-operation between some forty public and ten academic libraries and with support from the DALM is run as an open and free service for everybody.

In 1998 the work on a new Act on library services was initiated. The vision for the legislative work was to establish the framework for library services based on the needs of users in the information society. The Act was passed by Parliament in 2000.[2] It made access to the Internet and establishment of homepages obligatory for all public libraries. It stated that access to all published information – regardless of the medium on which it was stored – was free for any user. This led to a general building of collections of compact discs and multimedia materials in all libraries.

The Act stated a number of tasks that were the responsibility of the state. A major service was public access to a web-based version of the Danish Union Catalogue, which was launched in 2000 on 'bibliotek.dk' (English version: 'library.dk) containing all material registered in any public or academic library in Denmark, and giving access to search and request. Requested titles can be ordered to be picked up at any library chosen by the user. The success of this service convinced the DALM of the necessity to establish a national service for distribution of interlibrary loans. This national delivery service now delivers materials to libraries on a day to day basis, changing the policies for collection building and maintenance of many libraries. The library.dk portal is under constant development and improvement. In 2009 we plan to add user generated tags and ratings and we have agreed with the biggest Danish publisher, Gyldendal, to integrate access to the biggest Danish encyclopedia in library.dk. This integration will result in a search facility where you will be offered access to the full text article by subject or name searching.

From 2000 to 2005 the strategy of DALM was to support new web-based services in libraries, and many portals and subject-gateways were established. Among the most successful are an internet-guide, an e-zine on fiction, another on music and a portal for information for immigrants on the Danish society in eleven languages ranging from legislative to practical local community information covering the whole country. Other examples are e-support for home work.[3]

These services are provided by libraries in close co-operation. Majority of public libraries in Denmark contribute to this kind of web-based services and see them as integrated part of the new library concept.

From 2005 DALM started developing a new strategy for library development. One of the first elements analysed was the web-services of the public libraries. Several problems were identified. Among the major challenges were the lack of co-ordination and a coherent plan for development of these services on a national scale, the lack of clear criteria for success, and the cost-effectiveness of the services. A new organizational model has

[2] Act regarding library services. Copenhagen 2001

[3] Some of the library-driven webservices and portals can be found at the list in 'bibliotek.dk', the union catalogue: <http://bibliotek.dk/netbib.php>

resulted in a central co-ordination of the development and financing of the services. Plans and priorities are decided by a board representing the municipalities and the board. This step towards a more co-ordinated development has in the first place lead to a decision to establish a new library website for children. This website is planned to interact with daily children's library telecast that are planned from the autumn 2009 as well as with activities in the children's library space. Such an effort to create 'cross-over' from one platform to another and letting virtual and real offers to support each other would probably be one of the trends in the media-development in the coming years.[4]

In 2005 a new service was presented giving access to remote download of music files from the users own devices on a 24/7 basis via the library that pays for the license. This service was a major achievement in the effort to fight music piracy by supplying an alternative. This is particularly important in a small country like Denmark, where music piracy is a severe threat to the Danish music scene.[5]

The first step in the project was a digitization project carried out by The State and University Library in Aarhus that serves as national music collection. All published Danish music was digitized and an agreement with the association of right holders (an association of this kind is rare but probably a prerequisite for a broad national license) negotiated by a consortia consisting of DALM and a number of central libraries.

Initially the service was limited to Danish music, but international music is now available as well. At the moment the license covers more than 1 000 000 tracks, that can be downloaded by library patrons in municipalities covered by the license agreement. Almost all munipalities offer this service. The files can be borrowed for a day or a week (free of charge for the users). A new addition to the agreement will make it possible for libraries to give access for users to listen to 30 seconds of any track in the database without payment from the library. A similar service for download of films and videos from your home computer via a local library license has been launched in June 2009.

After a number of years where libraries have been aiming at giving access to e-books and other e-texts and have tested various models a new service has been launched, e-bog.dk , in a cooperation between the Association of Publishers and The Danish Bibliographic Centre. The service aims at e-access to all examinations requirement texts in all Danish education. Publishers have been somewhat reluctant in agreeing on e-services, as they are not yet convinced that the business-models running internationally will work in a national context as small as the Danish. DALM is constantly pushing the publishers to accept the conditions of information accessibility in the knowledge society. Now a service is running but at rather slow speed.

A new public library service moves faster. This service gives access to downloading of e-talking books to ipods and a number of other devices. The services was inspired by the same service offered by the Danish Library for the Blind.[6] But has now proved to be a very popular service for the general public. The major problem with this service is the high price the libraries pay per download.

[4] The strategy 'From Information to Knowledge' is only available in Danish, 'Fra information til viden' Copenhagen 2005

[5] Access to the services can be established at any local library in Denmark. See link: http://netmusik.shop2download.com/cgi-bin/WebObjects/TShop.woa/wa/PSHelp/faq

[6] The Danish Library for the Blind changed in June 2009 its name to 'Nota' in consequence of the fact, that the biggest user group are not blind people but dyslexcia users.

Also e-access to films seems to be a field with great potential. Libraries have for some years offered schools and other educational institutions access to web-based download of film with presentation in a DVD-quality. The titles are made available by cooperation with the Danish Film Institute and a few other rightholders. Now the general public also have access to this library service. The collection is still limited, but growing with more than a hundred titles per year.

Since the year 2000 where The Act Regarding Library Services first established the equality of status of the different types of media, the libraries have from many users experienced a quick shift from physical media to web-based delivery of the same information. Still the use of physical media exceed the digital, but the user pattern is clear: digital use is growing analog use is shrinking. The strategy in the public library area has subsequently been to establish remote accessible e-services in all fields and to develop more personalized and interactive services in the library place. This is an ongoing process.

DEFF – brief overview

Denmark's Electronic Research Library was established as a project in 1997 following an analysis of the development of a range of research libraries into electronic research libraries. The idea was to establish 'Denmark's Electronic Research Library', via a network of cooperating electronic research libraries. The initiative was launched as a five year project by three ministries with a total funding of approximately 27 million euros over a period of 4 years from 1998 to 2002.

By the end of the project period in 2002 the ministries decided to make DEFF as a permanent organization with its own funding on the National Budget. Since 2003 DEFF has functioned as an organizational collaboration between educational and research libraries in Denmark with an annual budget of 2.3 million euros.

In the period from the 2003–2006 the activities were organized around six focus areas: e-learning, e-publishing, licenses, portals, system architecture and user facilities. The primary objective of all activities was the development of local library service. It was thus highlighted in the DEFF strategy that the local library is the primary gateway for the servicing of end users. In 2006 the thematic focus areas were reduced to three: meeting the user, architecture and middleware and information provision. This division reflects DEFF's three-tiered architecture as the overall architectural blueprint for digital library development. The three layers are: joint information resources at the lowest level, middleware and webservices provide functionality at the intermediate level, and the user interface ensures the correct presentation for the user. It is a service oriented architecture where the three layers are independent of each other. The model of a three-tiered architecture functions both as a technical and organizational concept, because it allows the individual institution to collaborate on joint functionality and procurement of information resources, but preserves its own institutional user interface in interaction with the end user.

DEFF's new organization is supposed to ensure a stronger focus and a better coordination between the activities. It has, however, also been an opportunity to discuss more general aspects of digital library development and to re-examine DEFF's overall vision and the strategy for realising it. The focus areas retained the focus on local library development.

The priority of the development of local library services within DEFF is closely linked to the way IT-competence and development resources are distributed within the organization. IT development is carried out in

collaboration with employees at the individual libraries in projects that receive funding from DEFF. The original justification for this model was to ensure that solutions developed in one library are available to a wider range of libraries thus minimizing parallel development. The local institutional perspective was furthermore seen as a way to increase the chances of user-oriented solutions and to ensure that the projects were relevant to the libraries' specific needs, in this way increasing the possibility of the projects being actually implemented and sustained after the development phase.

The increasing complexity of IT development and the need for professional solutions are putting increasing pressure on this model of IT development because many of the smaller libraries are unable to maintain proper IT development departments and thus contribute to the development of IT solutions. Consequently, there is an increased focus on consolidation of IT development. These consolidation efforts are also extended to the operation of IT systems at the individual institutions because it is proving very difficult to develop IT solutions for a lot of different IT platforms. This continues to be an issue for DEFF.[7]

[7] The development of DEFF can be studied in the three reviews of DEFF:

i) Denmark's Electronic Research Library, Review of Denmark's Electronic Research Library 2008, Copenhagen 2008. Details available at <http://deff.intrasuite.dk/showfile.aspx?IdGuid={E231456E-1548-4406-9EE0-22D470B4EEE2}>

ii) Denmark's Electronic Research Library, Plans and perspectives for Denmark's Electronic Research Library – An international review, Copenhagen 2005. Details available at <http://www.deff.dk/showfile.aspx?IdGuid={C1613369-ECD5-46DA-A0FD-623F0E2336A9}

iii) Denmark's Electronic Research Library, Developing Denmark's Electronic Research Library - An International Review, Copenhagen 2001 . Details available at <http://www.deflink.dk/upload/doc_filer/doc_alle/806_DEF%20Review%20report%20.pdf

Developing digital library services – a general perspective

Developments in the library's environment

Analysing the changing information environment in a Danish context is probably no different from the patterns and trends identified elsewhere. The changes identified are primarily related to the digital information environment and the changed user behaviour it has caused.

The shift from 'collections to connections' and the changes in the information environment from a situation of information scarcity to information overload have together with increased use of search engines created a new breed of self-sufficient users who do not see the library as the centre of their information environment. These developments have created new working conditions for all libraries. The library collection only constitutes a small part of the relevant supply of information, and the library faces increased competition from new sources of information. The end result is an increased risk of 'library bypass'. It is no longer necessary to visit the library in order to get the information you need.

These developments have been recognized and discussed for many years, but very few libraries have taken the challenges seriously and taken steps to actively formulate strategies to counter these developments. We have had a lot Danish user surveys, usability studies, web statistics, and behavioural studies which indicate that libraries are increasingly losing their traditional strong ties with their patrons.

OCLC report – criticism and recommendations

OCLC's Pattern Recognition Scan from 2003 provided a very useful overview of these tendencies, and it also succeeded in highlighting some of the underlying causes.

The overall conclusion of this exhaustive analysis was that libraries are too focused on collections and that they should shift their focus to the end user. There was great emphasis on the term 'info sphere', which was defined as the network of information surrounding the individual.

The general recommendation was that the library's services should be brought to the user's 'info sphere' thus avoiding library bypass. Another interesting recommendation was that libraries should invest more resources in developing and offering collaborative technologies to users.

The increased popularity of collaborative technologies is seen as an expression of a more general trend of collaboration among users in the information environment. The most common term for this development is web 2.0 and the distributed collaboration this term implies. This trend means that users demand facilities for communication with other users and tools for further processing of information and for communicating the results.

The trend towards increased collaboration is one of three major trends highlighted in the conclusion of the analysis. The other two trends are disaggregation and a decrease in guided access to content.

Disaggregation means that services, institutions, information, and other entities in the digital environment are increasingly split into smaller parts and made available as such. There is for instance a demand that functionality must be available independently of the particular service. This concept is an important part of a service oriented infrastructure. But the trend is quite general and can be witnessed when a music CD is split into individual tracks, a book sold in separate chapters or an institution is viewed as a collection of services rather than a coherent entity.

The decrease in guided access to content is a natural consequence of the development and better quality of search engines. This trend implies that link collections are redundant and that libraries should not try to select and present only those information resources that are best suited for the user. The user has apparently become self-sufficient with the help of search engines. This means that information must be searchable and be presented in multiple contexts outside the library.

Although the report focuses on these three trends it points to a development that could be seen as a reaction or a complement to the trends: the demand for context. The exact meaning of the term is not defined but it points to the user's need for a frame of reference or assistance in a situation where: 'the people and institutions that acted as guides to content disappeared into a virtual world and have not been replaced in any meaningful way'.[8] The replacement of guided access to content with increased use of search engines is thus accompanied by a need for assistance and context at the point where the search result is presented to the user. The libraries' use of link resolvers (for example, SFX) and recommendation systems can help to create such a context and provide access to relevant resources at the search result. In the physical library context can be provided by supplementing the widespread introduction of self-service with the possibility of social interaction around the information resources.

Changes in education and learning

The reduced significance of traditional institutions in a digital environment could be seen as a strengthened individualization of services in the sense that the individual end

[8] <http://www.oclc.org/reports/escan/future/collaboration.htm> from OCLC Online Computer Library Center, Inc., The 2003 OCLC Environmental Scan: Pattern Recognition report, Alane Wilson (ed.) Details available at <http://www.oclc.org/reports/escan>

user has been given more choices. It would however be wrong to view the end user as the isolated individual or traditional customer of commercial companies, because the emphasis is increasingly on collaboration and sharing of information. This can be seen very clearly by observing these tendencies in education and research.

The education sector is experiencing the same tendencies as the libraries. There is a shift in focus from education to learning (decrease in guided access to content) and learning is taking place in the 'info sphere' and not on the blackboard. Learning is increasingly a collective activity (collaboration) and it is not based on specific textbooks or institutions, but combined from multiple courses and information sources.

It is no longer instruction in specific skills taking place at a particular institution but facilitation in order to promote the development of meta-competencies for lifelong learning. These concepts have been used as buzz words for many years, but they have been given a new significance because the competitive pressure of globalization and outsourcing on the job market means that the traditional education system is inadequate for handling these new challenges. The rate of outsourcing and creation of new, more skilled jobs mean that there is a huge demand for constant learning and development of new competencies. The traditional Danish education system is neither flexible nor large enough to meet this demand for further education. People will increasingly have to learn together outside a traditional institutional setting. This could be a good news for libraries.

Science and research are also experiencing developments that are challenging traditional institutional settings. E-science and mega-science are making the traditional collaborative nature of science even more pronounced and underlining the need for new institutional collaborations. Traditional scientific institutions are also losing their monopoly on research as both government and private sector research is becoming increasingly important. Of particular relevance to libraries are the new forms of scientific publication and communication and the focus on the availability of those datasets that form the basis of scientific publications.

Consequences for library development

These developments are changing the role of the library and have actually in the Danish discussions prompted some people to doubt whether there is a future for a digital library. The argument is that search engines, commercial information providers and the Internet have rendered the digital library redundant.

The fast changing nature of the information environment does indeed make it very important to continuously discuss the libraries' tasks and mission. There is a tendency in these discussions to equate traditional ways of doing things with the library's core competencies. This is highlighted by Lorcan Dempsey in the opening quote above. He urges librarians to focus on the mission instead of particular activities. He also formulated those activities in a very general way for instance using the term 'organising' instead of 'cataloguing'.

Nevertheless it still seems relevant to ask whether we need a particular institution to select, organize, retrieve, and transmit information when the Internet and search engines are developing the way they are. We would argue that there is indeed a need for such an institution, but that libraries generally need to shift their focus and think more strategically.

We agree with Lorcan Dempsey that focus should be shifted from particular processes and activities to a broad formulation of the

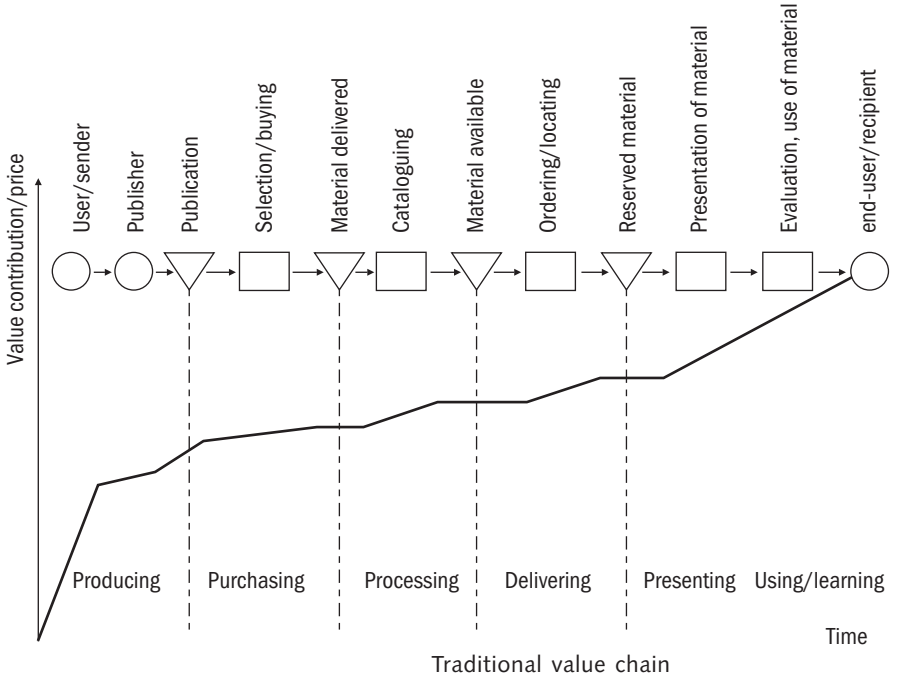
library's mission. We think that it would be more useful to replace the notion of supporting learning as the library's mission with the notion of facilitating the creation of knowledge. This is perhaps better suited to cover the library's role in relation to research and to underline the fact that the end user is not merely an information consumer but an information producer as well. Many libraries are thus taking on a new role in regard to electronic publishing.

A definition of the libraries role as facilitating the creation of knowledge implies a more pro-active role for the library than merely providing information. It is a very ambitious role in a digital environment, and it is not entirely clear what the implications might be.

In order to retain and develop the more traditional role of information provision the library services must be present in the user's info sphere which means that libraries must either develop services that are sufficiently visible and effective to attract users, or they must make the library's information resources available via the user's preferred commercial search engines.

The library's own resources in terms of printed and electronic collections only constitute a small part of the potential supply of information relevant to the user. It could be argued that as the information supply on the Internet expands and search engines become better, information provision is fast becoming a commodity service. If this is the case it should encourage libraries to standardize and consolidate this task and focus on other services where the library can add more value.

In developing such services the term 'user focus' should be taken quite literally, because it represents a shift in focus not only from the individual institution but also from the traditional task of information provision to the user's needs. This would imply a shift in focus from supply to demand and an effort to gain a much better understanding of the demands of the individual user.



Traditional value chain

An intuitive analysis of the information value chain from author/producer to reader/recipient would thus indicate that the largest increase in value takes places at the beginning and the end of the value chain. The model above (on page 77) is a simplified presentation of the different elements in the traditional value chain from the information producer (author) to the information receiver (reader). The model is intended to illustrate that the creation of value is largest at the beginning and end of the chain.

If the library is to add value to an institution or an end user, it should focus on the services at both ends of the value chain and consider outsourcing or consolidating the intermediate processes in the information provision. This is in fact what most libraries have done as we will try to argue later. If the library's role is indeed to facilitate knowledge creation, then helping the user or the institution acquire or produce knowledge could be the starting point for developing new services.

This would suggest following important focus areas for libraries.

- Teaching and publishing (institutional level)
- Communication and information literacy - (individual level)

The activities for the digital library in the first point are of course concerned with e-publishing and e-learning. The second point covers the library's role in enabling students to function as communicating participants in the knowledge society. This requires both information literacy (and i-skills) and the ability to express oneself and communicate using some of the many web 2.0 communication facilities.

Developing a new role for the library

The shift in service concept from 'what you see is what you get' to 'what you need is what you get' could help alleviate the risk of library bypass.

This requires the library to devote more resources to user studies, and user involvement including developing a range of services utilising web 2.0 functionality.

Web 2.0 has two major advantages from a library point of view. As it transforms communication from being bilateral (user-library) to being multilateral (user-user-library) it saves resources for instance in relation to reference services as users can answer each others questions and share advice on good information resources. The other very important advantage is that it gives libraries' a unique opportunity to analyse and understand users' information searching behaviour making it easier for libraries to continuously develop relevant library services.

Information provision

The very traditional role of information provision has undergone significant changes. The libraries have recognised that many of the processes involved, although central to the library role, does not add particular value to the end result. In line with traditional commercial wisdom for such processes the libraries have outsourced, consolidated and developed partnerships.

Many libraries have outsourced selection, acquisition and processing of materials to vendors. In Denmark the cataloguing of materials are shared on a national level or outsourced to vendors. DEFF handles negotiation and administration of electronic subscriptions and licenses. Much of the IT administration or development is outsourced to vendors or other libraries.

The processes surrounding lending are to a large extent outsourced via self-service to users. This includes ordering and reserving books, checking out and returning books and in some cases even being directly involved in the acquisition of books. We are fast

approaching a situation where the users can check out and return the books in a library without staff involvement. Information provision in the form of printed books is one area where the library's traditional role of information provision has indeed become a commodity service.

Digital library services

There is a tendency for digital libraries to focus on the task of information provision not merely an as important task, but as the task that justifies the library's existence. This often means that attempts to consolidate or standardize the task are viewed as threatening to the individual institution. It is considered very important for the individual institution to be visible to the end user, both in the role of information provision and via the user interface. This sometimes makes it difficult to collaborate on developing new services in this area.

Fortunately Google and Amazon have introduced a new standard for digital information provision that has increased libraries' motivation to cooperate in more areas. Two of the most important drivers for cooperation and consolidation in information technology are cost and complexity. Libraries have now realised that they have neither the skills nor the resources to deliver the functionality that users have come to expect from Amazon and Google – unless they cooperate.

Libraries have also come to recognize that users expect to be able to access all available information from one search box and to get exactly the result they need because of excellent search technology. Most libraries' information provision has traditionally been centred on the library catalogue, and this has been the starting point for the ambition to deliver a Google-like search interface. For many libraries this ambition demanded that all the digital information resources was catalogued and made available through the library catalogue. This strategy proved insufficient, because the library system is ill suited to handle the task, the amount of digital information being too huge, and because of the manual labour required.

An alternative approach is using federated search. This allows users to search multiple information services form a single search box but it normally utilizes the search facilities of the individual services which means that it is difficult to provide an intelligent ranking and presentation of the combined result.

The current approach under consideration in Denmark is to use integrated search. This demands that data in full text or metadata be aggregated and indexed in order to provide the same intelligent search and presentation to the users as Google delivers.

This approach has the advantage of presenting the user with very relevant search results but many of the information resources will not necessarily be immediately available to the user because the library's own information resources only constitute a very small part of the information presented. The presentation of so many information resources external to the library makes automated document delivery services and complex AAA (authorization, authentication and accounting) natural components of integrated search.

In the Danish context libraries are discussing the establishment of a mutual national data repository to be used as basis for the individual libraries' indexing and searching solutions. Other important areas of cooperation include AAA, document delivery, data on user behaviour and webservices to be used in adding relevant information to the presentation of the search result.

Concrete activities and service development

The different aspects of access control embodied in AAA are becoming increasingly

important, because the transition from print to digital means that 'fair use' is replaced by contractual regulation of the use of digital material. The ability to control and limit the use of information resources is often a precondition of making the material available in the first place. There is a need for a standardized and unified system across libraries and vendors which preserves the privacy of the end user while at the same time enabling single sign on to many different information resources. In a Danish context the project working with Shibboleth is seen as a step towards such a solution. The current initiative for access management for Danish research and higher education is called WAYF (where are you from?). By identifying the user's primary institutional affiliation the system makes it possible for the user to be authenticated at the parent institution regardless of which service the user wants to access. This means that user data is only maintained at one place thus increasing efficiency, security and privacy in the overall system of access management.

Comprehensive and user friendly solutions for integrated search, automated and flexible systems for document delivery and a common AAA-system are important components of an efficient information provision. The functionality required for collaborative filtering, recommendation services, customization and customerization is fast becoming a minimum requirement for the user interfaces to information resources and DEFF is looking to develop this functionality as services in the common architecture. The DEFF support in regard to library development of integrated search is primarily concentrated on building a joint data repository 'a data well' containing abstracts, metadata and full text from scientific journals and e-books. This data well will form the foundation upon which libraries will implement integrated search systems like Primo developed by Ex Libris or the locally developed open source system Summa.

These activities are supplemented by cooperation with Google Scholar with the objective of making the Danish union catalogue and other relevant Danish digital information resources available through Google scholar, thus increasing the visibility of these resources to the end user. A separate focus area in a Danish context is the digitization of Danish scientific and educational material. Legal difficulties and reluctance from publishers have so far blocked attempts to digitize these materials on a larger scale, but initiatives like Google Print and the EU's i-2010 initiative have contributed to a gradual opening in this area.

Institutional repositories and open access

DEFF's activities in the area of E-publishing started with projects on institutional repositories and the migration of journals to open access. Inspired by the e-framework proposed by JISC discussions are continuously taking place as to how the systems established would support research registration (Current Research Information System) and how the underlying repositories can interact with those who support e-learning. There is an increased focus on library support for the general knowledge management for the institution in all the processes between import and export of information. These deliberations reflect the more general idea that the libraries take on a much more active role in supporting the information flow and knowledge management for institutions and individual users.

An important driver in this development is the need for systems not only to handle publications and preprints but the underlying data sets and part of the research communication as developments within e-science and e-research seem to suggest. Some libraries are expanding their activities

into new areas such as developing research statistics and offering citation analysis and benchmarking of research activity. In addition to these demands in relation to research there is a demand for repositories for student papers and assignments.

The Danish libraries and DEFF have been very successful in cooperating on the development of CRIS (Current Research Information Systems). The collaboration has both included development of a common data model, an organizational structure to support collaboration and the adoption of identical systems in most institutions. This successful collaboration is a useful foundation for the further development of the infrastructure for promoting research registration and open access in Denmark.

A very important part of the work in this area is the technical and legal framework for the activities. Danish political awareness of open access has been very modest, but new initiatives are emerging and there is a Danish franchise of the creative commons movement. At the institutional level we are witnessing the formulation of guidelines for publications demanding that researchers deposit copies of preprints in institutional repositories.

At the national level The Ministry of Science, Technology and Innovation is introducing a metrics based funding system for research. The system is developed in order to ensure that increased future funding for research is allocated to the most productive research environments. The funding system will utilize data from the Danish Research Database which aggregates data from the CRIS at the individual university. In that way the DEFF-supported system of research registration is being developed to support a central metrics-based allocation of research funds. The implications of the metrics-based research evaluation system might be beneficial to the libraries' role as information providers if it turns out that there is a need for full text in order to judge allocation of points awarded to individual research articles. Such a need would encourage researchers to deposit copies of publications in institutional repositories thereby making it easier for libraries to make the publications available in support of open access. On the other hand the system might contribute to the preservation of the existing publishing system insofar as only subscription-based journals are included.

Organizational aspects – institutional partnerships

There is probably not one right way to develop the library at a particular institution. The local libraries develop according to local competencies and demands from the parent institution. The more conceptual discussions of the role of the library could nevertheless prove important, because it helps alleviate fears of the imminent demise of the library and tries to point to new tasks for the library.

These discussions might also clear the way for more rational cooperation between libraries and a better division of labour between the local library and national digital library services. In Denmark there are discussions as to whether libraries could divide their services into back office and front office functions, thus allowing the further development of shared services and outsourcing. These discussions are particularly important in the area of information provision where the concept of a local digital library is being challenged by increasing technical complexity and user demands.

The concept of the library role as being one of facilitating knowledge creation is sufficiently broad to allow individual libraries to develop services in new areas where they can probably add more value than trying to develop the traditional role of information provision on their own. The new tasks would allow the libraries to develop a new and closer

relationship to teachers, students and researchers. Such a role would imply a closer integration with the parent institution where the library would be perceived less as a monolithic institution and more as a network of services. In order to fulfil this role it would be important for libraries to cooperate nationally or internationally in the area of information provision and to collaborate in developing the new services and competencies in new areas. As an organization of research libraries, DEFF sees these discussions as essential to the continued development of the individual library and as a very important contribution to deciding which digital library services to develop on a national level.

Book review

Information Management in New Millennium: opportunities & challenges for library professionals

Sahu, Ashok Kumar

Ess Ess Publications, New Delhi
2008. 521 pages
Hard Bound
Price Rs 1250

*Reviewed by Dr Rabindra Kumar Mahapatra**

World Digital Libraries 2(1): 83

Due to revolution in technology, the information services have increased their variety and quantity and refined their quality. Information management is a crucial job for the library professionals in the twenty-first century. The professionals are now being challenged by new environment which demands changes in their styles, attitude and skills towards organizing and processing of information. Information technology has also tremendously changed the role of library to information centre and librarian to information manager. This book is very helpful to library professionals in understanding the management of information in new environments. This a practical book.

This is an edited volume with a compilation of 29 papers in different areas of information management. The contributions to this boom are from professionals and scholars from eminent institutions from India and abroad. This book has been edited by Mr Ashok Kumar Sahu. He has 14 years of professional experience in various libraries and has written several papers for national/international journals and conferences.

The book is divided into four parts- Information and Society, Information Management, Digital Library and Electronic Resources, and Future Trends. The papers in the first part discusses the impact of information on society and how the present society depends on the information. The second part discusses about managing, marketing of information and quality management in libraries. The third part explains the concept of digital library, how to digitalize documents, RFID security systems, and electronic resources. The fourth part of the book explains about future generation of library and expert systems.

The book explains how to manage, organize, market, the concept of digital library. It is a valuable resource for professionals, librarians, academicians and students of library science.

* Reader, Department of Library Science, Sanjay Memorial Institute of Technology, Berhampur, Orissa

News

World Digital Libraries 2(1): 85–86

India's Traditional Knowledge Digital Library: a powerful tool for patent examiners

On 2 February 2009 the Indian government granted access to its TKDL (Traditional Knowledge Digital Library), a unique database that houses the country's traditional medical wisdom, to examiners at the EPO (European Patent Office). EPO examiners will use the extensive database to prevent attempts at patenting existing traditional knowledge, a practice described as 'bio-piracy'. The co-operation between India and the EPO comes at a time when many countries are struggling to protect traditional and respected knowledge against exploitation, primarily in the pharmaceutical sector. Experts at the EPO say that access to the 30-million-page database will help to correctly examine patent applications relating to traditional knowledge.

UNESCO launches digital library at http://www.wdl.org/en/

UNESCO launched the WDL (World Digital Library) on 21 April 2009. This global digital library will provide a one-stop shop for researchers, teachers and schoolchildren seeking to find items on one topic at one place. WDL offers an invaluable platform for the free flow of information, for international solidarity, for the celebration of cultural diversity and for the building of inclusive knowledge societies. The digital library is to provide a free, easy-to-navigate site with a collection of primary documents on all subjects and authoritative explanations from the world's leading libraries. It brings together cultural heritage that's scattered around the world. The WDL supports UNESCO's mission of building capacity in developing countries, and intends working with UNESCO partners and external funders to establish digital conversion centres worldwide. These centres will produce content for the library as well as for other national and international initiatives.

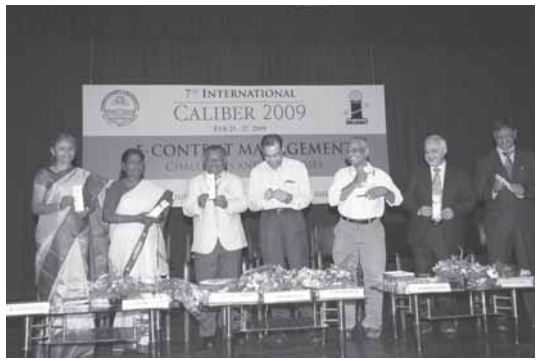
Chemical Abstracts Service launches free online database

The CAS (Chemical Abstracts Service) has launched Common Chemistry, a free online database containing information on 7800 chemicals of widespread and general interest as well as all 118 elements from the periodic table. Common Chemistry is primarily intended for the general public, including students, teachers, and the media. The database can be searched by chemical name or CAS registry number. Entries show a

compound's Registry Number, molecular formula, and chemical structure; chemical names and synonyms; and a Wikipedia link if available, McCue says. The Common Chemistry database will be updated periodically and further Wikipedia links will be added.

Maharashtra gets its first digital library
Maharashtra's first digital library with modern facilities will be unveiled in two weeks. It is the first full-fledged library being set up in the city after the State Central Library was established in 1947 in Town Hall at Fort, Andheri (E). The library will cater to students from schools and colleges, and those preparing for competitive exams for professional courses. The library is expected to cost Rs 7 crore of which Rs 1.5 crore will go towards infrastructure. Among other things, an MoU has been signed with DELNET (developing library network) to enable resource sharing with over 1500 libraries in India and other nations in South Asia. There is a proposal to connect it via Internet to the British library and the University Grants Commission library.

Seventh International CALIBER-2009
25–27 February 2009, Pondicherry University, Puduchery
The 7th International CALIBER-2009 was organized by the INFLIBNET Centre, Ahmedabad and Pondicherry University



during 25–27 February 2009. The main theme were e-content management, e-resources, Web 2.0, Library 2.0, content aggregation and application of ICT in libraries. The author presentations were grouped into four different sub-themes and were presented in four technical sessions spread over three days. The convention received 240 papers and after careful screening, 70 papers besides 3 theme papers and 110 abstracts were selected for publication in the proceedings. A total of 275 delegates have registered for the conference. An exhibition was also hosted during the event. At the end, the experts present in the conference, put forward the need for a national digital preservation policy in India. The committees set up by the Ministry of Information and Communication Technology would prepare a white paper on the subject and draw implementation strategies from various stakeholders.

Forthcoming events

World Digital Libraries 2(1): 87–88

Second International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2009)
4–6 August 2009, London, United Kingdom
Details available at <http://www.dirf.org/diwt2009/index.asp>

World Library and Information Congress: 75th IFLA General Conference and Council: Libraries create futures: Building on cultural heritage
23–27 August 2009, Milan, Italy
Details available at <http://www.ifla.org/annual-conference/ifla75/>

The International Conference for Digital Libraries and the Semantic Web 2009 (ICSD2009)
8–11 September 2009, Trento, Italy
Details available at <http://www.icsd-conference.org/node/4>

Second International Conference on the Theory of Information Retrieval
10–11 September 2009, Cambridge, United Kingdom
Details available at <http://kmi.open.ac.uk/events/ictir09/>

13th European Conference on Digital Libraries
27 September–2 October 2009, Corfu, Greece
Details available at <http://www.ecdl2009.eu/

International Conference on Preservation of Digital Objects (iPRES 2009)
5–6 October 2009, San Francisco, California, USA
Details available at <http://www.cdlib.org/iPres/ipres2009.html

International Conference on Academic Libraries (ICAL–2009)
5–8 October 2009, New Delhi, India
Details available at <http://crl.du.ac.in/ical09/

Fourth International Conference on Digital Information Management (ICDIM 2009)
1–4 November 2009, Ann Arbor, Michigan, USA
Details available at <http://www.icdim.org/>

IADIS International Conference: Cognition and Exploratory Learning in Digital Age (CELDA 2009)
20–22 November 2009, Rome, Italy
Details available at <http://www.celda-conf.org/>

Cultural Heritage online: Empowering users: an active role for user communities
15–16 December 2009, Florence, Italy
Details available at <http://www.rinascimento-digitale.it/index.php?SEZ=545>

Guide to authors

World Digital Libraries is an international peer-reviewed biannual journal. The journal seeks quality research papers that present original theoretical approaches. It also seeks experimental case studies related to digital library developments, maintenance, and dissemination of digital information focusing on research and integration of knowledge at the interface of resources and development. The journal will, therefore, keep readers abreast with the current developments and contain articles, reviews, current developments, and case studies, encompassing the following areas.

- Theoretical and methodological issues that relate to the interrelationships among electronic resources management, digital preservation, multiple access, multilinguality, copyright issues, and security aspects.
- Theoretical approaches as well as experimental case studies related to digital library development and maintenance.
- Initiatives towards digitization through lucid case studies.
- Current developments across the globe.
- Dialogues between the scientific community and society at large.

Articles should examine concepts, analyses, and case studies of important issues in the field.

Book reviews should be of recent publications in the field, to be reviewed by an independent reviewer.

Commentaries should discuss critical issues in the field.

Submissions

Authors are requested to send a soft copy (in Microsoft Word format) of their contribution to the editor, either in a CD or as an e-mail attachment.

All submissions will be peer-reviewed using the criteria of originality, accuracy, and quality of contribution in these fields.

Presentation of manuscripts

Articles must be original, in English, and should not exceed 8000 words. The main text should be double-spaced with headings and sub-headings clearly indicated in the text. All tables, figures, and equations should be numbered in Arabic numerals and clearly cited in the text. All measurements should be in metric (SI) units. The manuscript should be arranged in the order given below.

- Short title (10 words is the desired maximum length), subtitle (if desired)
- Author's name, affiliation, full postal address, and e-mail, telephone, and fax numbers (respective affiliations and addresses for co-authors should be clearly indicated)
- Abstract (not exceeding 200 words)
- Main body of the text, suitably divided under headings
- Acknowledgements, if any
- References
- Appendices (each on a separate sheet)
- Tables (each on a separate sheet)
- Figures (each on a separate sheet)

Shorter items

The following shorter items are also welcome and must be typed in the same way as major articles.

- Commentaries (research notes and short communications) and case studies (maximum 5000 words)
- Book reviews (maximum 1200 words)

In-house style: references

In the text, the surname of the author(s) followed by the year of publication of the reference should be given, for example, (Hall 1993). In case of several publications by the one author or by a group of author(s) in one one year, use notations '1993a', '1993b', and so on. Up to three authors can be mentioned in text references; more than three authors should be limited to the first three authors' names followed

by '*et al*'. References must be listed alphabetically at the end of the paper (double spaced) and should conform to the following style.

For journals

Davis G R. 1990
Energy for planet earth
Scientific American **263**(3): 55–62

For books

Carmichael J B and Strzepek K M. 1987
Industrial Water Use and Treatment Practices
London: Cassell Tycooly. 291 pp.

For chapters of edited books

Sintak Y. 1992
Models and projections of energy use in the Soviet Union
In *International Energy Economics*, pp. 1–53 edited by T Steiner
London: Chapman and Hall. 350 pp.

For grey literature

Togeby M and Jacobsen U. 1996
How conflicting goals concerning environment and transport influence the policy process?
Paper presented at the *Conference on Transport, Energy and Environment*,
3–4 October, Helsingor, Denmark

WBCSD (World Business Council for Sustainable Development) and UNEP (United Nations Environment Programme). 1998
Industry, fresh water, and sustainable development
Details available at <www.gm-unccd.org/FIELD/Private/WBCSD/freshwater.pdf>, last accessed on 9 January 2004

Footnotes

Authors are requested to use as few footnotes as possible, and keep their length to the minimum. Footnotes should be indicated in the text by superior Arabic numerals, which run consecutively through the paper. They should be grouped in order of appearance at the bottom of the concerned page in numerical order and must be double-spaced.

Accepted manuscripts

On acceptance, contributors are requested to provide the editor the final version of the article in soft and hard copy. Please observe the following instructions.

- Tables, figures, illustrations, should be on separate sheets.
- Retain a back-up disc for reference and safety.

Proofs

One set of proofs will be sent to the author before publication, which should be returned promptly within 48 hours of receipt. Authors are urged to check the proofs carefully as late corrections cannot be accepted.

Offprints

Apart from one free copy of the journal to the authors, 10 free offprints will be supplied to the first author. Further offprints and copies of the journal can be purchased at a reasonable cost, if ordered when sending the final copy of the article, or when returning the proofs.

Copyright

The responsibility for the contents of the paper rests with the authors, not with the editor or the publisher. Contributions are accepted for refereeing on the understanding that they have been submitted only to this journal and not to any other journal. Only when each author signs and submits the CTA (Copyright Transfer Agreement) can TERI Press publish the article. This CTA enables TERI Press to protect the copyright material for the authors, but does not affect the authors' proprietary rights. The CTA covers the exclusive rights to reproduce and distribute the articles, including reprints, photographic reproductions, or any other reproductions of similar nature and translations, and includes the right to adapt the article for use in conjunction with computer systems and programmes, including reproduction or publication in machine-readable form and incorporation into retrieval systems. Authors are responsible for obtaining, from the copyright holder, permission to reproduce any figures for which copyright exists.

Acknowledgement

The editor thanks all those have graciously given their time to review papers for *World Digital Libraries* throughout 2008. The efforts, much appreciated, have contributed to the improvement and value to the articles and the journal as a whole.

Reviewers of the *World Digital Libraries: An International Journal* **Volume 1**, 2008

Prof. José Luis Borbinha, INESC-ID/IST, Information Systems Group, Lisbon Technical University, Portugal

Dr H S Chopra, Head, Bhai Gurdas Library, Guru Nanak Dev University, Amritsar

Prof. Gobinda Chowdhury, PhD, Director, Information and Knowledge Management, University of Technology, Sydney, Australia

Dr S M Dhawan, Library and Information Consultant, Former Scientist F, National Physical Laboratory, New Delhi, India

Prof J N Gautam PhD, Department of Library and Information Science, School of Social Science, Jiwaji University, India,

Prof. Alan Hopkinson, Head of Library Systems, Learning Resources, The Sheppard Library, Middlesex University, UK

H Anil Kumar, Librarian and Head NICMAN, Indian Institute of Management, Ahmedabad, India

Prof. Ee-Peng Lim, PhD, School of Information Systems, Singapore Management University, Singapore

Dr Madhusudhan Margam, Lecturer, Department of Library and Information Science, University of Delhi, Delhi, India

Prof. Gary Marchionini, School of Information and Library Science, University of North Carolina at Chapel Hill, USA

Dr S Majumdar, University Librarian, Central Library, DULS, University of Delhi, Delhi, India

Dr Usha Mujoo-Munshi, Librarian, Indian Institute of Public Administration, India

Prof. Paul Nieuwenhuysen, Vrije Universiteit Brussel, Belgium

Prof. H N Prasad, Head, Department of Library and Information Science, Banaras Hindu University, Varanasi, India

Prof. (Dr) K C Sahoo, Head of the Department, Department of Library And Information Science, Dr Harisingh Gour University, Sagar, India

Dr Ratna Sanyal, Indian Institute of Information Technology, Allahabad, India

Dr M P Satija, Prof and Head, Department of Library and Information Science, Guru Nanak Dev University, Amritsar, India

Prof. Dr Michael Seadle, Director, Berlin School of Library and Information Science, Humboldt University Berlin, Germany

Gleb Skobeltsyn, LSIR, Distributed Information Systems Laboratory, Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland

Prof. M G Sreekumar, PhD, Librarian, Indian Institute of Management, Kozhikode, India

Prof. Shigeo Sugimoto, University of Library and Information Science, Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan

Prof. P Tapio Varis, PhD, Acting President, Global University System, Unesco Chair in global eLearning with applications to multiple domains, Professor and Chair of Media Education, University of Tampere, Finland

Prof. Om Vikas, PhD, Former Director, ABV Indian Institute of Information Technology and Management, India

World Digital Libraries: an international journal

World Digital Libraries is an international peer-reviewed biannual journal. The journal seeks quality research papers that present original theoretical approaches. It also seeks experimental case studies related to digital library developments, maintenance, and dissemination of digital information focusing on research and integration of knowledge at the interface of resources and development. The journal will, therefore, keep readers abreast with the current developments and contain articles, reviews, current developments, and case studies, encompassing the following areas.

- Theoretical and methodological issues that relate to the interrelationships among electronic resources management, digital preservation, multiple access, multilinguality, copyright issues, and security aspects.
- Theoretical approaches as well as experimental case studies related to digital library development and maintenance.
- Initiatives towards digitization through lucid case studies.
- Current developments across the globe.
- Dialogues between the scientific community and society at large.

How often is WDL published?
WDL is published twice a year, in June and December. The two issues make up a volume.

How to subscribe and how much does it cost?
Annual subscription to WDL (print + online) is Rs 1500 within India and overseas subscription is $150. Payment can be made by demand draft or local cheque drawn in favour of TERI payable at New Delhi. One can also subscribe through TERI's online bookstore <http://bookstore.teriin.org>.

Whom to contact for more information?
For information on editorial content and for submitting manuscripts and books for review, please contact The Editor, WDL. For more information about the journal log on to <http://bookstore.teriin.org/journal_inside.php?material_id=477&qty=1>

For enquiries about subscriptions and advertisements, please contact
TERI Press
The Energy and Resources Institute
Darbari Seth Block
IHC Complex
Lodhi Road
New Delhi – 110 003

E-mail teripress@teri.res.in
Web www.teriin.org
Fax 2468 2144 or 2468 2145
Tel. 2468 2100 or 4150 4900
India +91 • Delhi (0) 11

Guide to authors
See pp. 89–90

International Conference on Digital Libraries (ICDL)

Shaping the Information Paradigm

New Delhi • 23–26 February 2010

Venue: Conference at India Habitat Centre, New Delhi. 24–26 February 2010
Tutorial at IGNOU, Convention Centre, New Delhi. 23 February 2010

The Third International Conference on Digital Libraries (ICDL) is being organized by TERI in partnership with IGNOU, during 23–26 February 2010. The theme of the Conference is *Shaping the Information Paradigm*. The conference will focus on creation, adoption, implementation and utilization of digital libraries, e-learning and a knowledge society. The conference aims to provide an international forum for sharing of experiences among researchers, educators, practitioners, and policy makers from a variety of disciplines such as library and information science, information and communication technology, archival and museum studies, knowledge management and many areas in the fields of social sciences and humanities.

Conference highlights

- 50 internationally renowned speakers will share their views.
- Awards will be given to the '*best paper*' and '*best poster*'.
- Selected papers will be published in the WDL journal.
- International vendors and publishers will exhibit their products.

Call for papers

Original papers focusing on the theme of the conference—Digital Libraries: Shaping the Information Paradigm are invited for the conference poster and tutorial.

Listed below are some of the topics.

- DL development, architecture, and management
- Contents management in DL
- Multi-linguality and interoperability issues
- Digital rights management
- Digital preservation and access management
- Semantic web
- KM (knowledge management) and organizational repositories
- E-learning and e-publishing
- DL standards and policy
- Open archives initiatives
- ODOL (Open distance online learning)
- Multimedia content
- Virtual support to distance learners
- E-resources management for distance learners
- Access management
- m-learning technology

For details of paper submission guidelines and submission process, visit <www.teriin.org/events/icdl>.

Contact

Debal C Kar
Organizing Secretary
ICDL Secretariat
TERI, Darbari Seth Block
IHC Complex, Lodhi Road
New Delhi – 110 003, India

Phone 24682138, 24682100, 41504900
Fax 24682144, 24682145
E-mail icdl@teri.res.in
Web www.teriin.org/events/icdl
India +91 • Delhi (0)11