Technological Undercurrents and Global Information Circulation.

Ian Pettman
Pettman & Associates
P.O.Box 30
Grange-over-Sands
Cumbria LA11 7GS
UK

ABSTRACT: Briefly reviews some of the wide range of technological developments arising from and driven by the Internet. Focuses on explaining the developments in librarianship terms and indicates some opportunities offered to library and information workers through these developments. Discusses how some of these technologies have been used in some aquatic Internet information services and how some developments might be used in the near future. Uses examples from several aquatic and fisheries information services such as *oneFish* and *Freshwater Life.*

KEYWORDS: Information services; information scientists; Internet; subject portals; Z39.50; XML; RDF; taxonomies; ontologies; classification; metadata; thesauri

Introduction

Although this session of the Conference is actually meant to cover Internet Systems and Tools in the sense of "services," I thought it might be useful to try and give a fairly basic overview of the *underlying* systems and tools which will be driving the new Internet portals and services. Whilst doing this, I hope that there will be many areas where you can see opportunities for us, as library professionals, to assist with the further development of these as well as how we can use them in order to achieve that primary library objective Peter Brophy articulated as "To enable users to gain access to the information they need."

Any Internet site has the potential to contribute to effective global information circulation. There are certain challenges – many of which will be very familiar to us as librarians. However, this familiarity can sometimes be obscured by the terminology used and I hope that I may also be able to assist with overcoming this barrier for those of you who have been bemused by the unnecessary techno-babble.

The Aquatic Sciences Libraries and Information Centres (IAMSLIC. EURASLIC, BIASLIC etc – the Aquatic SLICs) have been making wide use of the Internet with their email; discussion groups; and websites for many years now. These services are constantly being refined and further developed e.g. the latest IAMSLIC and EURASLIC sites at www.iamslic.org and www.euraslic.org .

The Aquatic SLICs are also trying to bring both some order and added value to the library side using the Z39.50 initiative. EURASLIC was involved in some of the early work in developing multiple catalogue searching and Inter Library Loan using Z39.50 – as partners in the European Union UNIverse Project. IAMSLIC is now actively utilizing Z39.50 using the NOAA sponsored gateway to make multiple Aquatic SLIC catalogues simultaneously searchable over the web. http://www.csc.noaa.gov/CID/iamslic.

As John Akeroyd said, in his Keynote Paper to this Conference, librarians have been at the forefront in the uptake of the Z39.50 standard. There are some problems in applying Z39.50 to heterogeneous systems such as library catalogues. These can be and are being overcome and libraries will continue to use this standard for some time to come. However, although he did not elaborate the point, John also said that he thought Z39.50 might give way to other standards or systems, possibly based on XML. As this paper progresses, it should be possible to see how this may come about.

> ***What is Z39.50?*** Z39.50 is a communications standard which describes the rules and procedures for communicating between two computer systems for searching and retrieving information from databases. Most library system vendors have incorporated the standard into their cataloguing and other applications. By using one of these systems, the searcher is able to search many different databases accessible through Z39.50 servers and retrieve results. The searcher uses the same search interface for all searches and does not require a detailed knowledge of the system being searched.

However, with the rapid growth in availability of full text documents, other media (especially data) and a wide range of technological developments (the ***technological undercurrents*** of the title) there are further opportunities for libraries and librarians both to bring order and to add value to this broader field of information resources and activities.

Technological Undercurrents – an overview from a librarian's viewpoint

The last 3 to 5 years have seen some tremendous innovations in relation to the underlying technologies of the Internet. The terms and general concepts are often heard at Conferences such as this, but it is not always easy to see how these fit together and how we might take advantage of them. The terminology can confuse and discourage us.

Consider a possible example – you and your colleagues may have decided to pursue another of John Akeroyd's points and decide to become more active in exploiting the available global knowledge base. You have a good idea for an Internet global information service or Gateway (similar to oneFish [1] or FreshwaterLife [2] say)– it looks big – you don't feel confident - you go out to a specialist company – you get a proposal something like the following:

"The Information Architect will analyse your proposed Gateway and produce a report that will detail the Vortal strategy; Markup Language options; interoperability requirements; metadata standards. The report will also provide an analysis of the ontologies and taxonomies arising from the subject content to underpin the browsing, navigation and searching systems and present you with a technical blueprint from which the website development team will be able to build your site's infrastructure.

To complete this stage there would be a one-off charge of: $45,000"

This is a real world example – not a construct just for this paper! Intimidating isn't it? How does this design guru – the Information Architect - cope with this complexity? No wonder she can charge such sums of money. What do all these technological undercurrent words mean?

If we extract the jargon terms and phrases from this proposal:

Vortal strategy
Markup languages
Interoperability requirements
Metadata standards
Ontologies and Taxonomies

they all pivot around the central essential concept for a global information system – INTEROPERABILITY. Because of both the quantity and the transitory nature of Internet information, no global information service will be sustainable if it cannot handle distributed data effectively. In order to do this, every aspect of systems design and all the underlying technologies have to be centred on interoperability.

Vortal Strategy - Gateways, Portals and Vortals

As we all know, web based information resources are growing exponentially. Gateways, Portals and Vortals are all approaches to try and introduce ***selection policy*** into/onto Internet information resources.

Koch [1] has reviewed definitions and suggested typologies for gateways that are useful, not least in showing the differences that exist between broadly similar services. Working definitions will, however, do for our purposes:

Gateways provide search services to other people's web resources - ***Subject gateways*** provide search services to other people's web resources selected from a particular subject area - ***'Quality controlled subject gateways'*** make quality assessments before including resources.

The word ***Portal*** means gateway and sometimes is synonymous but portals often offer more services than gateways i.e. provide more than just a search service to other people's *web* resources. The first web portals were online services, such as AOL, that provided access to the web plus services, such as e-mail, forums, search engines, and on-line shopping malls. Now most of the traditional search engines have transformed themselves into web portals to attract and keep a larger audience.

Specific subject Portals, ***Vortals*** (Vertical portals), offer extended services into their specific subject areas. Vortals typically provide news, research resources (documents etc), data, discussions, newsletters, online tools, and many other services that educate users about a specific subject. oneFish is an example of a Vortal.

oneFish is an Internet vortal for fisheries and aquatic resources research which adopts a participatory approach, allowing users to submit knowledge as well as discover it – stakeholder selection policy. [http://www.onefish.org/]

FreshwaterLife is a proposed distributed web vortal co-ordinating data and information on the taxonomy and ecology of freshwater plants and animals in Europe and North America. It is also envisaged that it will enable data analysis and community building. [http://www.freshwaterlife.org]

Any worthwhile vortal, like its special library counterpart, and with regard to the distributed nature of its resources, needs a constantly evaluated ***Collection Development Policy*** and ***Service Definition Document*** as well as a wide range of ***Consortium Agreements***– or in Internet-speak, a vortal strategy.

Opportunities for librarians

- Involvement with Collection Development Policy
- Advising on Service Definition
- Negotiating Consortium Agreements for Information Quality and Provision

Since we are considering a service with distributed content, there obviously has to be interoperability between the data on the various content holder sites – i.e. standards for the way information resources are structured, cataloguing standards, and standards for data manipulation. This brings us to Markup Languages, Metadata Standards and Resource Description Framework.

Markup Languages and Interoperability Requirements

Up to three years ago, any Internet site would have been based its content and display on Hypertext Markup Language (HTML) and the consideration of other markup languages would not have been necessary. This all changed in 1998/99 when a new language was released - XML. What is it, and why has there been so much excitement about eXtensible Markup Language?

Markup Languages

HTML

The web's main language is (or was) HTML. Although HTML is the most successful electronic-publishing language ever invented, it is superficial. In essence, it describes how a web browser should arrange text, images, etc on a page. HTML's concern with appearances makes it relatively easy to learn, but it also has its costs.

One major cost is the difficulty in creating a web site that functions as more than just a fancy fax machine sending documents to anyone who asks. People want web sites that take information from distributed sources (and sometimes users), transmit and manipulate both textual and data records, even run scientific instruments from half a world away. HTML was never designed for such tasks.

HTML uses tags to tell the computer about layout. The solution to getting more interactive websites is, in theory, very simple - use tags that say what the information is, not what it looks like.

XML and the "X-Files"

Extensible Markup Language (XML) is a new language designed to do just that, to make information self-describing. XML lets everyone create her own tags to annotate web documents for meaning e.g <Favourite malt whisky> or <Guin Auction Item>. Thus XML–defined web pages can function like database records.

This simple-sounding change in how computers communicate has the potential to extend the Internet beyond just information delivery to information manipulation and analysis and has been taken up very rapidly.

Another source of XML's unifying strength is its reliance on a still fairly recent standard called Unicode, a character-encoding system that supports intermingling of text in all the world's major languages. Thus, XML enables exchange of information not only between different computer systems but also across national and cultural boundaries.

Unlike most computer data formats, XML markup also makes sense to humans, because it consists of nothing more than ordinary text.

Of course, it is not quite that simple. Although XML does allow anyone to design a new, custom-built language to describe his information resources, if we want to "interoperate", we have to get groups of interested people to concentrate on agreeing exactly how they want to represent the information they commonly exchange (cf the development of MARC).

Groups of people have, however, gotten together and developed agreed "Activity-specific interchange languages". Indeed, a shower of new acronyms ending in "ML" testifies to the inventiveness unleashed by XML in the sciences, in business and in the scholarly disciplines [see Applications section below].

Before they can draft a new XML language, these groups must agree on three things:
which tags will be allowed,
how tagged elements may nest within one another and
how they should be processed.

The first two--the language's vocabulary and structure--are typically codified in a Document Type Definition, (DTD) or Schema.

XML DTDs or Schemas then enable data exchange - but XML tags offer no inherent clues about how the information should look on screen or on paper. So now we have excellent data interoperability but no display. We now need to apply rules organized into "style-sheets" to reformat the work automatically for various devices so that the information can be displayed or rendered.

XSL
The standard for XML style-sheets is called the Extensible Stylesheet Language, or XSL.

The ***latest*** versions of several web browsers can read an XML document, fetch the appropriate XSL style-sheet, and use it to sort and format the information on the screen. (or even as audible speech or a Braille print out or a tune if the text was music). However, for users with earlier versions of browsers, the XML has to be formatted by XSL to HTML. At the present time, therefore, XML is mainly used as a server side language. It has been so successful and generated such excitement because ***it is a simple, standard way to interchange structured textual data between computer programmes*** – essential for interoperability.

XSL formats the output into the required format such as HTML, PDF or Comma-Separated Values but a first step is often needed – that of ***transforming*** the structure of the incoming XML document to a structure that reflects the desired output.

XSLT
XSL Transformations (XSLT) can be used to transform one form of XML into another form of XML. XSLT is a powerful language for transforming XML data in many ways. These may involve selecting data, grouping it, sorting it, or performing arithmetic conversions.
An example would be to take an XML format for monthly sales figures and produce a histogram as its XML output using the XML-based SVG standard for vector graphics. A more leisure-based example would be to take a piece of musical notation (ChordML) and, using XSLT, generate a Musical Instrument Digital Interface (MIDI) file and play the music on a synthesizer.

With this scope for transformation, XSLT also enables and empowers interoperability.

XLink
One of the web's main innovations was "hypertext," its billions of pages connected by hyperlinks--those underlined words you click on to get whisked from one page to the next. Hyperlinks, too, will do more when powered by XML. A standard for XML-based hypertext, named XLink, allows you to choose from a list of multiple destinations and, perhaps most useful, enables authors to use indirect links that point to actual entries in databases.

XPath
XML Path Language (XPath) seems to have taken over from XML-QL as the preferred form of query language within XML – allowing you to perform queries on XML documents and data [the equivalent of SQL (Standard Query Language) for relational databases].

So, if you need extensibility (user defined tags) and interoperability, it makes sense for you to use XML as part of the basis for your site development. If you do, which subject and activity specific interchange languages (the applications) should you consider?

The Applications – the MLs

XML is a mother tongue for other languages, so application markup languages to suit your specific needs become possible. There is so much activity in this area that whatever subject or activities you are considering for your vortal there may already be a working group and possibly a draft ML. To find out, one initial port of call would be http://xml.coverpages.org/sgml-xml.html Under "Applications", the site has several hundred entries under "XML: Proposed Applications and Industry Initiatives". Examples include:

VocML - Vocabulary Markup Language
this will support the structured representation of authority files, hierarchical thesauri (including those with poly-hierarchies), classification schemes, digital gazetteers, and subject heading lists.

BSML - Bioinformatic Sequence Markup Language
The proposed Bioinformatic Sequence Markup Language (BSML) is a public domain protocol for Graphic Genomic Displays. The project goals are in some respects similar to those of the Chemical Markup Language.

IDML - International Development Markup Language
IDML would become a data exchange standard for information that is specific to international development, making it much easier to share information with regional offices, partner agencies and with the public.

These MLs can describe either ***resources*** or ***activities*** and give us the ***standards*** enabling us to interchange or inter-lend our globally distributed information stock.

Many ***metadata standards*** are now also being expressed in XML.

Metadata and Metadata Standards: "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information." [Hodge [2]] Traditional library cataloging is a form of metadata, and MARC 21 and the rule sets used with it such as AACR2 are metadata standards.

Dublin Core is probably the most widely known metadata standard but there are very many more e.g.–

OAMS - Open Archives Metadata Set
- presents a technical and organizational framework designed to facilitate the discovery of content stored in distributed e-print archives.

EAD – the Encoded Archival Description

Vortals need catalogue data in order to work efficiently just like any library.

Dublin Core:
In 1995, 52 researchers and practitioners concerned with libraries and networking gathered in Dublin, Ohio, to attempt to arrive at a list of descriptive metadata elements intended to promote ***author generated*** resource description for web based documents. It was thought that Internet authors might use a simple element set of about 15 elements (Title, Creator etc). This became the **Dublin Core** element set. It has been translated into more than 15 languages, is in use in more than 50 projects, and is the basis for describing official documents in at least 2 countries. On 5th October 2001, the Dublin Core Metadata Element Set was approved by ANSI as a recognised standard **(Z39.85-2001)**.

Internet documents, pages or resources do not necessarily have detailed metadata (a catalogue record). This makes it difficult to get accurate search results, increases search time, and leads to several other problems and inefficiencies. The uptake of standards, such as Dublin Core, by Internet authors is still slow but does seem to be growing.

Extensions and Profiles

Despite the recent development of many of these metadata schemes, most have already been subject to the changes brought about by implementing them in real world situations. These modifications are of two types: ***extensions*** and ***profiles***.
An ***extension*** is the addition of elements to an already developed scheme to support the description of an information resource of a particular type or subject or to meet the needs of a particular interest group. Extensions increase the number of elements.

Profiles are subsets of a scheme that are implemented by a particular interest group; profiles can constrain the number of elements that will be used, refine element definitions to describe the specific types of resources more accurately and specify values that an element can take.

Although the original objective of the Dublin Core was to define a set of elements that could be used by ***authors*** to describe their own web resources, the library community, particularly our digital library colleagues, have begun the process of developing a Library Application Profile for Dublin Core. A draft version is available at http://dublincore.org/documents/2001/08/08/library-application-profile/

> **Profiles**
> This has also happened with Z39.50. As the number of libraries using Z39.50 increased, some difficulties became apparent – e.g. the user receiving many false hits or, conversely, not retrieving a record even if it was in the database. These problems lead to the introduction of Z39.50 Profiles. ***The Bath Profile**: An international Z39.50 specification for library applications and resource discovery* identifies those features of the Z39.50 standard that are required to support effective use of Z39.50 software for a range of library functions, such as basic searching and retrieval of bibliographic records for cataloguing, interlibrary loan, reference, and acquisitions.

Another interest group working to develop an application profile of Dublin Core is The Agstandards Discussion Group. This is a group of information management specialists in the domain of agriculture that discusses various issues on agricultural information management. In June 2001 a draft **Metadata set for the description of agricultural documents and document-like resources** was issued for discussion and trailing as a potential application profile of Dublin Core
http://www.fao.org/agris/MagazineArchive/magazine/TaskForceonDCMI.htm.

As stated previously, most of these metadata standards are now expressed in XML – and some pre-XML metadata standards have been "translated" into XML e.g.

BiblioML - XML for UNIMARC Bibliographic Records

However, what happens if you want to incorporate elements from more than one metadata standard in your resource description? From the outset, part of the XML project has been to create a sister standard for metadata that also provides a mechanism for integrating multiple metadata schemes – this standard is called the **Resource Description Framework (RDF).**

Example 1: RDF and Multiple Application Schemas

```
<? xml version="1.0" ?>
<RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
 xmlns:DC = "http://purl.org/DC#"
 xmlns:AGLS = "http://naa.gov.au/AGLS#" >

 <Description about = "http://dstc.com.au/report.html" >
 <DC:Title> The Future of Metadata </DC:Title>
 <DC:Creator> Jacky Crystal </DC:Creator>
 <DC:Date> 1998-01-01 </DC:Date>
 <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
 <AGLS:Document> Instructional </AGLS:Document>
 <AGLS:Function> Information Management - Internet </AGLS:Function>
 </Description>
</RDF>
```

The first line of Example 1 simply indicates that this is an XML document. The next three lines indicate three namespaces – RDF as the default namespace, Dublin Core and the Australian Government Locator Service (AGLS) metadata schema.

The main section of Example 1 - between the <Description> tags - shows six Properties that describe the resource pointed to by the URL <http://dstc.com.au/report.html> in the *about* attribute in the <Description> tag. The first four properties come from the Dublin Core (DC) namespace (the Title, Creator, Date, and Subject) and the last two from AGLS (Document and Function).

RDF provides much more than this however. RDF gives meaning. RDF makes assertions in sets of triples. To take an example from the above RDF document – the web page **http://dstc.com.au/report.html** has property **"Creator"** with the value **"Jacky Crystal"**. This is a meaning that computers can "understand" and use.

oneFish is looking at both the Agricultural Dublin Core and IDML metadata sets to asses their use for fisheries and aquatic resources and possible implementation within the oneFish vortal. If oneFish uses multiple metadata sets in this way it will, of necessity, need to consider using RDF within its system design.

Pausing to take stock, where are we now – where have these undercurrents taken us so far?

We now have a set of resources and activities marked up so that they can be shared, transmitted, manipulated and displayed and we have metadata records (catalogue entries) for these resources. All of this is now interoperable between many global "hosts".

What we still need are systems to allow us to display them in logical groups (shelf order etc) for browsing; or to assist when searching. We need these systems to be capable of being used by many cultures and many language groups. This leads us on to our Information Architect's ***Ontologies*** and ***Taxonomies***.

The Internet community applies this terminology in varied and confusing ways leading to a lot of misunderstanding and confusion. However, in general terms, the two concepts cover: Classification schemes, Glossaries, Thesauri and Authority files.

Taxonomies

In many ways, Taxonomies could be considered to be part of the wider concept of Ontologies. In essence, they are a means of organising data for web sites to assist both browsing and search retrieval – sometimes classification schemes, sometimes thesauri and sometimes both. Some sites use more than one classification scheme (or even more than one thesaurus).

Opportunities for librarians:
The unbounded growth of Internet resources has given new prominence to the role of library classifications in imposing order on chaos. Any individual or organisation seeking to organise a large body of Internet resources is likely to call for a "Taxonomist" in the realm of knowledge management.

For a vortal, displaying the resources in a systematic topical arrangement is essential. Systematic arrangements can be created ad hoc, based on the collection content at a point in time, and then revised as the collection grows. Alternatively, they can be drawn from an existing Classification Scheme.

Many Internet sites are now using library classification schemes to represent their site's "taxonomy". DDC, UDC and LCSH are all extensively used on Internet sites. Some specialist schemes are also used e.g. the American Mathematical Society Classification.

Others use thesauri as their grouping structure e.g the EEVL and EELS engineering web sites both use subject categories based on the Ei Thesaurus (Engineering Index Thesaurus).

Ontologies

Put in its simplest form, an ontology is a collection of terms, definitions of the terms and defined relationships between the terms for a particular domain or subject area. This data set can then amalgamate the advantages of glossaries, classification schemes, thesauri, authority files (including biological taxonomy authority files) and gazetteers.

In other words, it can be thought of as an enhanced thesaurus—it provides all the basic relationships inherent in a thesaurus, plus it defines and enables the creation of more formal, more specific and more powerful relationships. An ontology captures and structures the knowledge in a domain (subject area), and by doing so captures the meaning of concepts that are specific to that domain. This meaning is then extended to end-users through the use of tools (e.g., indexing, retrieval and browsing tools) that apply the ontologies.

An ontology can be multi-lingual and, if several ontologies are properly constructed for different subject domains, they can assist ***cross-site*** and ***cross-subject*** searching and retrieval.

The following is a hypothetical example of what a simple ontology might produce for the term "Fish":

Fish as Noun [Verb variations not included]

Synonyms

Sense 1
fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
=> aquatic vertebrate -- (animal living wholly or chiefly in or on water)

Sense 2
fish -- (the flesh of fish used as food)
=> foodstuff, food product -- (a substance that can be used or prepared for use as food)

Sense 3
fish, chump, fool, gull, mark, patsy, fall guy, sucker, schlemiel, shlemiel, soft touch, mug -- (a person who is gullible and easy to take advantage of)
=> victim, dupe -- (a person who is tricked or swindled)

Sense 4
fish, go fish -- (a game for two players who try to assemble books of cards by asking the opponent for particular cards)
card game, cards -- (a game played with playing cards)

Hypernyms (Broader Terms)

4 senses of fish (only one displayed)

Sense 1
fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
=> aquatic vertebrate -- (animal living wholly or chiefly in or on water)
=> vertebrate, craniate -- (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
=> chordate -- (any animal of the phylum Chordata having a notochord or spinal column)
=> animal, animate being, beast, brute, creature, fauna -- (a living organism characterized by voluntary movement)
Etc.

Hyponyms (Narrower Terms)

2 of 4 senses of fish (only part of one displayed)

Sense 1
fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
=> cartilaginous fish, chondrichthian -- (fishes in which the skeleton may be calcified but not ossified)
=> holocephalan -- (fish with high compressed head and a body tapering off into a long tail)
=> chimaera -- (smooth-skinned deep-sea fish with a tapering body and long threadlike tail)
=> rabbitfish, Chimaera monstrosa -- (large European chimaera)
Etc

Holonyms (is part of; is a member of)

1 of 4 senses of fish (only part of 1 displayed)

Sense 1
fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
MEMBER OF: school, shoal -- (a large group of fish; "a school of small glittering fish swam by")
MEMBER OF: Pisces -- (a group of vertebrates comprising both cartilaginous and bony fishes and sometimes including the jawless vertebrates; not used technically)

MEMBER OF: Vertebrata, subphylum Vertebrata, Craniata, subphylum Craniata -- (fishes; amphibians; reptiles; birds; mammals)

Etc.

Meronums (has part)

1 of 4 senses of fish (only part of 1 displayed)

Sense 1
fish -- (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
HAS PART: fish scale -- (scale of the kind that covers the bodies of fish)
HAS PART: roe -- (eggs of female fish)
HAS PART: milt -- (seminal fluid produced by male fish)
HAS PART: lateral line, lateral line organ -- (sense organs of fish and amphibians; believed to detect pressure changes in the water)
Etc.

Example 2 - possible Ontology output [source: WordNet [3]]

This example shows the terms (or vocabulary); the definitions (or glossary); and some possible relationships given as a printed output.

We should note that sections 4 and 5 in this example are just two specific applications of the thesaurus concept "Related Term" – there are potentially many more such relationships.
We should also note that the example is only monolingual.

This type of printout from a hypothetical Ontology is only meant to give you some idea of the information that may be within the Ontology. As pointed out earlier, a Vortal would be designed to incorporate the Ontology tools within its system [usually in "background" mode] such that they assist the users with browsing, indexing and searching the Vortal's resources and tailoring them to their individual needs.

Terms (vocabulary) express concepts. Concept relationships are expressed in classification schemes. Classification schemes, therefore, also come under the umbrella of "ontologies".

At the present time there are many initiatives attempting to produce "cross-walks between different classification schemes.

Dewey, UDC, LC and Bliss are working together.

The IAMSLIC Coordinating Committee on Subject Analysis has been mapping ASFA descriptors against LC Subject Headings.

The Renardus Project [an EU project to establish a collaborative framework for European subject gateways [4]] is encouraging and enabling cross-classification mapping. Using DDC as its core system, Renardus is inviting participants to map DDC terms to their own "taxonomy". This includes DDC to UDC, Mathematics Subject Classification and Nederlandse Basisclassificatie.

The development of ontologies has progressed in several fields and particularly important work has been done for the medical knowledge sources. The US National Library of Medicine has been particularly active in this field with its Unified Medical Language System (UMLS [8]).

In the field of Agriculture and Fisheries, the UN Food and Agriculture Organization is leading an initiative for a proposed Agricultural Ontology Server project (AOS).

Although the AOS would use the AGROVOC thesaurus as a platform, **it will need to build associations with community partners** for its development. For instance, in the fisheries area, the AOS could partner with oneFish, ASFA and FIGIS (the FAO Fisheries Global Information System) among others. In the biological taxonomy areas, partners could include the Integrated Taxonomic Information System (ITIS [5]), the National Biological Information Infrastructure (NBII [6]), and the Federal Geographic Data Committee (FGDC [7]), again, among others.

Opportunities for Librarians:
Involvement with thesaurus term development – both mono- and multi-lingual
Involvement with the definition of more relationships between terms
Involvement with the classification cross-walk projects
Evaluation of the effectiveness of new software tools such as Topic Map software

The AOS is also looking at software tools to enable build, integration and maintenance of such an ontology. One software presently being evaluated for this purpose is Topic Map Software based on the 1999 ISO 13250 Standard for Topic Maps.

The Future

The rapid development of these technological undercurrents is showing no sign of slackening off. The phrase "Web Services" has gained visibility in the last few months –

particularly from the big industry players such as IBM, Microsoft and Sun. So, what is this all about and should we, as librarians, be involved?
Web services are interoperable building blocks for constructing web service applications. As an example, we can imagine a distributed library infrastructure built on web services providing functionality such as distributed search, authentication, inter-library loan requests, document delivery, document translation and payment for services – all tailored to particular user communities.

The industry giants have been working to develop a set of XML based open standards that enable the web service architecture to be implemented. There are 3 main components involved – WSDL, SOAP and UDDI.

The Web Service Description Language (WSDL) is a standard way of capturing service descriptions.
The Simple Object Access Protocol (SOAP) is a standard for XML based information exchange between distributed applications and hence communicates the service requested from the services described in the WSDL.
The Universal Discovery, Description and Integration (UDDI) is a specification for distributed registries of web services so that we know what web services are currently available.

These could be the XML building blocks that replace and enhance the Z39.50 standard we, as librarians, are presently using. Gardner [9] has given a useful introduction to this interesting development.

In the longer term, Machine Automated Indexing (MAI) and Natural Language Processing (NLP) will play a useful role in automated indexing and subject assignment. Visual Representation of Information Spaces will become more common for enhancing web browsing.

Conclusions

Global Information Circulation is becoming a reality and the Technological Undercurrents are essential components of the systems

This paper has attempted to explain some of these Technological Undercurrents driving the Internet information services in librarianship terms.

By trying to remove some of the techno-babble smoke-screen, I have over-simplified many of these developments – but I hope that I have at least shown that these are not impenetrable topics and that we, by our training and background in librarianship, have all the necessary skills to play valuable roles in progressing these developments.

At some future date, developments in Machine Aided Indexing, Natural Language Processing and Visual Representation of Information Spaces may displace the need for some of the librarianship skills – those that John Akeroyd, in his keynote paper, talked about abandoning anyway - but not just yet.

For the foreseeable future, other librarianship skills will be in demand for Internet information services – particularly in the areas of Markup Language standards, Taxonomies and Ontologies. Maybe it is possible for the title Information Architect to be considered as at least a partial synonym for Internet Librarian?

I would like to conclude then by paraphrasing another of John's keynote points –

"We have the skills, we have the adaptability, we have the high ground and, therefore, we have the possibility of being higher paid."

References

[1] oneFish Community Portal http://www.onefish.org

[2] Freshwater Life web site http://www.freshwaterlife.org

[3] Koch T. 2000. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review* 24 (1): 24-34.

[4] Hodge G., Metadata made simpler. www.techstreet.com/cgi-bin/detail?product_id=924055

[5] WordNet http://www.cogsci.princeton.edu/~wn

[6] Renardus http://www.renardus.org/

[7] http://www.itis.usda.gov/access.html

[8] http://www.nbii.gov/datainfo/metadata/

[9] http://www.fgdc.gov/standards/status/sub5_8.html

[10] http://www.nlm.nih.gov/research/umls/

[11] Gardner, T. 2001. An introduction to web services. *Ariadne*, Issue 29. [Online.] Available: http://www.ariadne.ac.uk/issue29/gardner/intro.html [June 10, 2002].