Future Library Systems :
Beyond the Electronic Card Catalogue

Geoffrey Payne
General Manager,
Information Services
Vision Australia Foundation
rmggp@ozemail.com.au

Abstract:
This paper contrasts the capabilities of traditional integrated library systems with the improving information access provided by the Internet. Recent standards developments are mentioned. Readers are urged to think about how a library needs to work to meet the expectations of users based on their information access experiences on the Internet.

Future Library Systems : Beyond the Electronic Card Catalogue

When computer based catalogues first appeared in libraries, users embraced them with enthusiasm, quickly abandoning the card catalogues which often still provided the only access to the bulk of libraries' collections. It took many years for large libraries to convert all their card catalogue information into computer based catalogues based on Machine Readable Cataloguing (MARC) records.

Today we are seeing screen based information replace printed information as the preferred medium for many purposes. The Internet as a focus for information delivery by governments, retailers, banks, and most organisations is well and truly upon us. The Internet is also a vehicle for delivering distance education and conducting electronic commerce.

While the way to incorporate the card catalogue into the early automated library systems was obvious, albeit resource intensive, it is not so obvious how we should integrate libraries and the Internet.

Traditional Integrated Library systems

The integrated library system had its origins in the control of physical library materials. The systems focussed on acquisitions, cataloguing, serials check-in, circulation of materials and materials booking.

Is this the business libraries should be focussed on today?

Cataloguing and Information Retrieval

The integrated library systems essentially automated libraries' card catalogues and serials check-in visible indexes. Minimal information is recorded about each item in the database: Author, Title, one or two controlled vocabulary subject headings assigned by a cataloguer, and a number from a classification scheme.

Thus a library catalogue can be great for "known item" searching, but poor at providing access to the totality of information in the collections.

Conventional Library System Constraints

Information in the item referred to by a catalogue record is largely inaccessible until the item is in hand. Chapter headings, paragraph headings, table of contents and index entries are not usually accessible, nor are they searchable from the integrated library system. Word or phrase searching of the full text of an item represented by a catalogue record is not possible. Article level access to periodical articles is often not available, or not available without authentication as a licensed user. There is no option to pay a fee in lieu of being a licensed user. Any audio or visual information about or included in the work catalogued or indexed is not available.

At present no standardised functionality applies to the enquiry functions of systems from different vendors. An author search may function differently when performed on systems from different suppliers. Thus when simultaneous searches are launched on several

catalogues by a Z39.50 compliant search engine, the enquiry is not necessarily consistently processed at each system, leading to uncertainty about the outcomes of the search process.

This is being addressed with the development of The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery.[i] This profile aims to standardise the functionality which must be supported by a Z39.50 server for searching a range of fields, and thus remove the element of uncertainty which is inherent in the results of searching multiple catalogues today. That this profile intends to standardise holdings searching as well as other catalogue record fields may mean we see some improvements in the mess that keeps holding statements human readable rather than machine interpretable today. This keeps inter-library loan transactions labour intensive, either for library staff or for users to whom we pass the problem of identifying a library holding an item in which they are interested.

Challenges Not Well Addressed by Integrated Library Systems

Library systems are typically poor for relevance ranking (as there is only the catalogue record on which to base the ranking), accessing the actual information catalogued, accessing the information content quickly. They are also poor at ensuring the information is accessible at all, as the object may be on-loan, lost, hidden or mis-shelved in the stacks.

I can download legislation and lots of other documents in all formats from the Internet, even some whole books. Why can't I do this for materials I would expect to find through my library service?

User Authentication

The OPAC is a public domain database owned by the Library, and for which user authentication is not required before access is granted. For databases licensed by Libraries on behalf of their clients, identification and authentication is required before access is granted. Simple originating IP address checking for incoming enquiries does not allow for students accessing the library's resources through their own Internet service provider, from their place of employment, or from their home domain if they are located offshore. Access needs to be based on a users' membership of a defined community rather than their network address. Applications proprietary to one vendor's system are less than ideal.

An overview of the technical solutions being explored at Columbia University illustrates the complexity of this issue.[ii] While some Universities may be able to invest the resources necessary to research and develop solutions, public, school and many university libraries need solutions "off the shelf".

Rights Management

In July 1998 the INDECS (Interoperability of Data in E-Commerce Systems) project technical co-ordinator, Godfrey Rust, criticised the Dublin Core metadata set[iii] as not containing elements necessary to support rights management. Dublin Core at that stage supported information discovery, but not the terms and conditions on which a copy could be accessed. Rights management itself is highly dependent on precise identification of individual items and hence is potentially reliant on the information and identifiers included in the Dublin Core metadata set. Rights management also presumes identifiers and authentication systems for users and rights owners. The Resource Description Framework

(RDF) developed by the World Wide Web Consortium enables the consistent encoding, exchange, and machine-processing of standardized metadata through exploiting the features of XML (Extensible Mark-up Language)[iv]. But where are the mechanisms for identification? Still in development or traversing the ISO standards processes?

Library systems typically lack rights management facilities. If one links content to a catalogue record there is no facility to track access and copying for charging purposes. There is no e-commerce module to aggregate large numbers of small transactions such as pay per view charges. This facility is desirable regardless of whether the Library or the end users pay the bills to the content provider. These constraints preclude use of electronic reserves or pay as you go texts unless royalties are waived or the library owns the copyright.

Where do libraries find identification, authentication, rights management and information hosting packages to allow us to meet copyright obligations and deliver on-line access to "library resources" without the need to visit the library?

On-line information on the World Wide Web

Despite the difficulties with user authentication and rights management, the Internet already has enough public domain information to have demonstrated that it is great mechanism for supporting:

> Find me something on …
> Full text on-line
> Keyword searching of full text
> Global searching through search engines
> Visual content on-line
> Audio content on-line
> Immediate access to information

But poor for "Known item" searching, as it is hard to refine searches based on the public search engines.

User Expectations of Electronic Service delivery

The Internet has spawned a push to electronic service delivery by governments and businesses as they seek to reduce the costs of providing services and make them more convenient for users, typically offering twenty-four hour access from anywhere around the world. Clearly the world has an expectation that the Internet, or at least a technology like it, is the public communications utility for the forseeable future.

This impetus is driving the rapid evolution of data formats and standards to allow all types of digital representations of information to be delivered over the Internet to a variety of browsing technologies. The common graphical web browsers most of us use today are already augmented with speech out browsers for the vision impaired. Soon accessing email and web based information from mobile telephones and other hand held devices will be commonplace. Thus audio, text only, limited graphics in black and white, will join the existing web browsers in rendering web based information for a variety of applications.

E-commerce applications such as Internet banking and shopping are also being rapidly developed, further embedding the Internet and preparing the ground for related technologies such as interactive television and datacasting in the mindset and behaviour of our communities of library users.

Extensible Mark-up Language (XML) and Cascading Style Sheets (CSS)

To ensure a single piece of information can be satisfactorily rendered by each of the above output mechanisms, Hypertext Mark-up Language (HTML) is being superseded by Extensible Mark-up Language (XML) as the basis for structuring information objects. XML enforces the separation of content and information on how it is to be displayed, the mixing of which in HTML is one of its greatest weaknesses. XML also contains sufficient features to permit a document marked up in XML to support the needs of publishers for producing fully featured print publications. This capability to render a single representation of an XML document in a variety of ways is achieved through Cascading Style Sheets (CSS)[v].

Cascading style sheets also allow a user to apply their own style sheet to override the display settings proposed by the document author. Thus a user can change colours if colour blindness is a problem, change font sizes to address low vision problems, and even specify intonations to be used for headings etc in an audio style sheet to assist in synthetic speech rendering of the data.

Synchronised Multimedia Integration Language

The SMIL standard defines a means of integrating text, video and audio into a multimedia presentation such as a television news clip. SMIL is not a replacement for the standards for the component parts (such as XML for Text, AIFF for audio and MPEG for video) but a mechanism for synchronising objects encoded in these formats into a presentation. SMIL does not describe the component parts of the presentation but how they are to be combined temporally and spatially to create a presentation.

SMIL has been developed by the World Wide Web Consortium Synchronized Multimedia Working Group.[vi] SMIL is intended to facilitate interoperability in this area.

How do XML compliant documents fit into library collections? Where does metadata beyond the MARC fields fit?

The Battle of the Search Protocols

Z39.50 has been developed as the protocol allowing integrated library systems to inter-operate. Z39.50 had its origins in the Linked Systems Project of the 1980's and is a complex protocol to implement. Z39.50 is still evolving, for exampl the development of the Bath Profile as described above.

The search for ways of implementing search capabilities in the Internet environment has seen various attempts to do something simpler than Z39.50. The Z39.50 Implementers Group commenced work on the Stanford Protocol for Internet Search and Retrieval in 1996. This project developed a simple protocol that text search engines should follow to facilitate searching and indexing multiple collections of text documents.[vii]

The ZDSR (Distributed Search and Ranked Retrieval) profile builds on the STARTS project within the Z39.50 protocol. ZDSR assumes that queries pertain to documents, and for each document there is a document descriptor consisting of metadata about the document including a pointer which may be used to retrieve the document; document retrieval is

otherwise out-of-scope of this profile. [viii] ZDSR was also influenced by a desire for simplicity.

These developments come down to paying a high price in development work for the richness of a full Z39.50 implementation or a high price in retrieval capability for a simpler retrieval mechanism. Hopefully the work that has gone into getting the Z39.50 standard to support complex enquiries will not be lost in the web environment. The challenges to Z39.50 seem to have waned in the last couple of years, but there are lots of other sectors developing web based services and exploring options for searching.

HTTP / Z39.50 Gateways

For the time being it is possible for users with a web browser as their client software to enjoy the benefits of Z39.50 through a Z39.50/HTTP gateway maintained by a third party such as a library. The gateway accepts incoming enquiries from the web browser, propagates them simultaneously to one or more target databases which support the Z39.50 protocol, the presents the result set to the web browser. The problem here is that most target databases contain metadata such as catalogue records, but not the documents themselves.

Simultaneous retrieval of information from multiple agencies

The pilot information service developed for the Zavier project aims to bring together the vast cultural information resources of six Victorian cultural agencies - the Museum of Victoria, the National Gallery of Victoria, the State Library of Victoria, the Public Record Office, Cinemedia, and the Victorian Arts Centre Trust through its Performing Arts Museum.[ix] Using Zavier a user could use their web browser to access a Z39.50 gateway to retrieve information from Z39.50 servers in each of these organisations.

Projects such as Zavier deserve wide promotion as illustrating what can be achieved using the functionality embodied in the Z39.50 standard. It is a pity the demonstration system seems to have disappeared. The links in the project report are broken.

Merging the Library and the Internet

With the Internet we have more focus on information management and access than ever before. We have a richness of information in text, graphical, video and audio formats, and whole documents routinely accessible rather than just metadata such as catalogue records or citations.

Where does this leave the integrated library system?
Should MARC records become just another document type in an XML world?
How can libraries meet user expectations in an environment increasingly focussed on instant access to documents?
Can we match the vibrancy of multimedia delivered on demand?
How can we retain the curriculum materials developed in web authoring tools?

We need to reinvent libraries, by keeping what we are good at, and adding value to what is out there, just as we have always done in the paper based world.

To this end, is it time to close off the integrated library system just as we closed off the card catalogues?

When we do, will a library service look like an organised corner of the Internet? Will libraries be members only clubs delivering licensed information and paying the utility bill at the end of the month? Will my Internet Service Provider be chosen based on its "library" service? Will it look like netLibrary, the on-line subscription library?[x] Will it comprise electronic books compliant with the open e-book standard?[xi]

Should we be loading XML tables of contents and indexes derived from publishers files for the physical materials we acquire rather than continue MARC cataloguing, or as an adjunct to MARC cataloguing. How do we get publishers to change from print only focus in their preparation of materials for publication?

Do we have a clear vision of what lies beyond the automated card catalogue? How can we drive technological developments to meet the needs of our users?

[i] The Bath Profile:An International Z39.50 Specification for Library Applications and Resource Discovery, Draft for Public Comment developed by The Bath Group (Participants at the Bath, UK Meeting, August 1999)
http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/bath/

[ii] Glenn, Ariel and David Millman. "Access Management of Web-based Services: An Incremental Approach to Cross-organizational Authentication and Authorization" in *D-Lib Magazine* (ISSN 1082-9873), September 1998
http://www.dlib.org/dlib/september98/millman/09millman.html

[iii] Rust, Godfrey. "Metadata: The Right Approach : An Integrated Model for Descriptive and Rights Metadata in E-commerce", in *D-Lib Magazine* July/August 1998 (ISSN 1082-9873)
http://www.dlib.org/dlib/july98/rust/07rust.html

[iv] Bearman, David et al. "A Common Model to Support Interoperable Metadata : Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities" in *D-Lib Magazine*, v.5(1) January 1999 (ISSN 1082-9873)
http://www.dlib.org/dlib/january99/bearman/01bearman.html

[v] Vanderheiden, Greg (ed). "WAI Page Authoring Accessibility Guidelines"
http://www.w3.org/TR/1998/WD-WAI-PAGEAUTH-0203.html#Style Sheets
& "Style Sheets Activity Statement" http://www.w3.org/Style/Activity.html

[vi] Synchronized Multimedia Integration Language (SMIL) Boston Specification W3C Working Draft 15 November 1999 http://www.w3.org/TR/1999/WD-smil-boston-19991115

[vii] STARTS Stanford Protocol Proposal for Internet Retrieval and Search http://www-db.stanford.edu/~gravano/starts_home.html

[viii] ZDSR Profile; Z39.50 Profile for Simple Distributed Search and Ranked Retrieval
http://lcweb.loc.gov/z3950/agency/profiles/zdsr.html

[ix] "ZAVIER Executive report summary"
http://dstc.edu.au/Research/Projects/Z3950/ZAVIER/ZAV-NewExecSum.doc

[x] "netLibrary offers the world's largest library of eBooks"
http://www.netlibrary.com/index.asp

[xi] "On September 21, 1999 the Open eBook Authoring Group released the final version of the new 1.0 specification" http://www.openebook.org/