

Metadata in Public Libraries



Walter Lewis

Halton Hills Public Library

The Halton Hills Public Library?

- ◆ Community of 50,000 people; 25 FTE including 3 professional librarians
- ◆ Fringe of the Greater Toronto Area
- ◆ Member of HALINET (HALton Information NETwork)
- ◆ Believers in the old notion of the “one place to look” (at least for accurate referral)

Metadata types

- ◆ Descriptive metadata
 - ◆ Resource discovery ...
- ◆ Administrative metadata
 - ◆ Creation, Rights ...
- ◆ Technical/structural metadata
 - ◆ Formats, location ...
- ◆ Collectively: Collection Management

Contexts for metadata in the Public Library

- ◆ Traditional library information systems
- ◆ Community information
- ◆ Local historical and genealogical information

Traditional Library metadata

- ◆ Collection catalogue
- ◆ Indexing services (magazines and newspapers)
- ◆ Marc centric services
- ◆ Purchased content and data management services

Community Information

- ◆ Directories of agencies and services
- ◆ Volunteer opportunities listings
- ◆ Events listings

- ◆ Developed web-centered software to facilitate collection, management and searching (<http://www.cioc.ca>)

Local Historical and Genealogical Information

- ◆ Newspaper
- ◆ Images (part of Images Canada)
- ◆ Books (including indexes of non-digitized material)
- ◆ Census, wills, property records, military records, cemetery, business directories, maps ...
- ◆ 600,000 + records, millions of searches/year

Maritime History of the Great Lakes

- ◆ Power of Partnerships: Entirely volunteer, contributed content
- ◆ Personal site and Development test bed
- ◆ Newspaper articles (TEI)
- ◆ Transcribed and imaged texts
- ◆ Images
- ◆ Enrolment and registry databases
- ◆ <http://www.hhpl.on.ca/GreatLakes/>

Data models vs metadata

- ◆ Single most important concept in presentation
- ◆ Metadata schemas are largely about how you encode content as you share it with other systems
- ◆ Good data models will allow you to support multiple metadata schemas
- ◆ The last thing you want to do is to derive a MARC or METS record from a database modeled to support Dublin Core

Examples (1)

- ◆ Images database:
 - ◆ Export routine to transform the data on its way to Images Canada (modified Dublin Core)
 - ◆ Working with one of the NCSA's Open Archives Initiative routines to support OAI (Dublin Core required; other transformations optional)
 - ◆ Crossnet's Zedlib templates to transform the same dataset into MARC and SUTRS to answer Z39.50 queries

Examples (2.1)

- ◆ Community Information database
 - ◆ CIOC's gateway (<http://gateway.cioc.ca/>) uses the Crossnet Zedlib toolset to deliver MARC 21 (Community Information) records via the DRA/SIRSI TAOS Web2 interface
 - ◆ Could as easily be integrated into a broadcast Z39.50 search oriented interface for a single library
 - ◆ Need to be able to pass address information off to mapping agencies (from the native web interface)

Examples (2.2)

- ◆ Community Information database (cont.)
 - ◆ Requirement to export to Microsoft Access for use in local derivative works.
 - ◆ Requirement to export and import from the ACICO standard format while engaged in other joint projects
 - ◆ Looming need for data exchange with the CCACs

Examples (3)

- ◆ Digital books using Text Encoding Initiative (TEI)
 - ◆ Shares with the EAD (and HTML) the notion of embedded metadata in a head area to describe the content in the body
 - ◆ Embed <index>, <name> tags in the body
 - ◆ XSLT to create web pages with Dublin Core meta tags
 - ◆ Export other tags to database to drive other kinds of searches
 - ◆ XML and databases are complementary, not mutually exclusive technologies

Description in binary objects

- ◆ Challenge of orphaned digital objects
- ◆ Separate work flows
 - ◆ Digitizers
 - ◆ Metadata specialists
 - ◆ ... often for excellent reasons
- ◆ Orphaned projects (and content)
- ◆ Projects that are working on this issue, but where the priority is rights and technical metadata

Resource discovery

- ◆ Native interfaces
- ◆ Google vs the Deep Web
 - ◆ There are strategies to expose this content to Google etc.
- ◆ Harvested content
- ◆ Broadcast search

Native interfaces

- ◆ Most projects have one
- ◆ Rarely closely linked to the metadata... schemas
...
- ◆ ... except when loading specific content into broader content management applications
 - ◆ e.g. the Milwaukee Public Library catalogue
 - ◆ (http://www.mpl.org/file/hum_marine_shipfile.htm)
- ◆ Maximum number of “rabbit holes” or “silos”
- ◆ Need to expose a crawlable interface to Google et al. – subject to their relevance criteria

Harvested content

- ◆ Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
 - ◆ Unqualified Dublin Core
 - ◆ Other metadata standards as configured
- ◆ Bi-lateral agreements
 - ◆ Images Canada (<http://www.imagescanada.ca/>)
- ◆ Directed web harvesting

Broadcast and Government Info

- ◆ Demand for simplified access to board range of government information
 - ◆ Municipal
 - ◆ Provincial
 - ◆ Federal
- ◆ Range of resource discovery options
 - ◆ Local, customized search interface
 - ◆ Support metadata and searching standards that allow organizations to define the interface, navigation elements that link the data

Broadcast searching

- ◆ Just-in-time searching
- ◆ One Place to Look
- ◆ Usually associated with Z39.50
- ◆ Can be built across a set of native databases
- ◆ Commercial applications (WebFeat, Agent etc.) combine techniques from screen scraping to being metadata aware

Parting thoughts

- ◆ The same data can support a wide range of metadata standards, as appropriate
- ◆ Customized local search interfaces still matter
- ◆ Standards-based interfaces that allow broader searching matter