Metadata Conversion and the Library OPAC

Amanda Xu
Serials Cataloger, Massachusetts Institute of Technology
Room 14E-210A, 77 Massachusetts Ave., Cambridge, MA 02139-4307, USA

Introduction

The number of scholarly information resources offered on the Internet has grown tremendously and will increase exponentially in the future. This development has pointed up the need for applying metadata to networked information resources, as it has been widely accepted that resources with metadata attached or embedded have the potential to be more easily discovered and retrieved via the Internet.

This paper examines the metadata movement on the Internet, and anticipates the need to establish a metadata repository for library collections. It suggests using the OPAC as a gateway to access the metadata repository, whether it is located on the library Website, or on the local databases mounted with specialized metadata formats.

The paper conceptualizes a metadata conversion system built into library OPACs for metadata integration and display. As a result, both locally created metadata and incoming metadata selected as part of the library collection can be captured via automatic metadata extraction, mapping, conversion, and integration. Consequently, the newly integrated metadata can utilize the full functionality of the library OPAC.

Metadata and the Metadata Movement on the Internet

Metadata, or representational data, is generally defined as data about data. It contains a set of data elements to describe the content and location of an information object and to facilitate its discovery and retrieval in the networked environment.

A library bibliographic record is one type of metadata. It differs from other metadata in that it uses *AACR* as rules for data modelling, USMARC for encoding schemes, and prioprietary online systems for information retrieval for more than three decades.

Most metadata on the Internet do not have such long-established and well-regulated standards, encoding schemes, and systems to support the mission of networked resources discovery and retrieval. But this is changing, and we have to examine the impact on the library world.

The Necessity for Metadata

Metadata is important in global information retrieval on the Internet for the following reasons:

- It manages large amounts of data with low network bandwith:
  Metadata addresses the issue of indexing large quantities of data of various types without requiring enormous amounts of network bandwidth. What gets indexed is the representational data rather than the information object itself.

- It assists effective discovery and retrieval of networked information resources:
  Metadata facilitates more sophisticated and comprehensive searching of information as the metadata elements and structures are designed to analyze content of the data in depth.
- It shares and integrates heterogeneous information resources:
  Information resources exist in different formats with different features residing in heterogeneous databases. Standard metadata description permits the comparing, sharing, integrating and reusing of various types of data in a distributed networking environment. Metadata has thus become an important long-term approach for finding information in heterogeneous databases.
- It controls restricted-access information:
  Metadata can not only facilitate effective searching, and retrieving of heterogeneous information resources, but it can also manage restrcited-access information and services to users, e.g. billing, filtering and rating, privacy, and security. Metadata serves a gatekeeper function, an indispensible feature for ever-growing commercialized information resources.

Metadata Formats and the Creation of Metadata

The movement to utilize metadata to organize the content of information on the Internet is expanding. Metadata can be created at the authoring stage, when the author embeds metadata into the resource using HTML META-tags, SGML headers, and/or other metadata templates. It can also be generated on the fly by search engines. For instance, metadata in SOIF (Summary Object Interchange Format) [1] is generated by Harvest gatherers. And it can be constructed at the service stage as well: the metadata is located in a central or distributed databases with pointers to the resources it describes.

Metadata formats have proliferated, ranging from simple and structured formats to rich formats, and moving from priopyrietary, emerging standards, to international standards. A comprehensive report documenting metadata description formats has been written by DESIRE (Development of a European Service for Information on Research and Education). [2]

Searching for Metadata

Not all resources with metadata attached will be discovered by search engines, because the types of metadata a search engine gathers depends largely on the types of metadata templates that are profiled. [3] For those "Internet accessible but non-HTML based resources," [4] metadata can be accessed via protocols such as Whois ++, [5] LDAP (Lightweight Directory Access Protocol), [6] Z39.50, [7] or other priopyrietary search engines. One such proprietary search engine is Panorama. [8] It is the primary software available for searching SGML-based encoding schemes such as the TEI Header [9] and EAD (Encoded Archival Description) [10] records. OCLC's Site Search, though it claims to be able to search SGML documents with a Z39.50-based search form, cannot effectively handle long text documents such as finding aids. [11]

Mapping Metadata

To improve metadata interoperability, metadata mapping has been seriously considered by the various metadata players. The most frequently mapped metadata formats are: IAFA (Internet Anonymous FTP Archive) templates, [12] Dublin Core metadata sets, [13] USMARC, GILS (Government Information Locator Services), [14] SGML TEI Header, EAD, [15] and Z39.50 tag set G. [16] Among them, USMARC, TEI Headers, EAD, GILS, and Dublin Core can represent the center of metadata mapping.

By mapping the content, syntax, and data elements of various metadata models, correct metadata conversion between various syntaxes can be assured. [17] Sketchy records such as IAFA and Dublin Core records can be accurately upgraded during the migration so as to satisfy the needs of rich description records such as USMARC, TEI Header, and EAD. Ultimately, searching across metadata syntaxes and databases can be facilitated.

The very fact that metadata can be mapped indicates that most metadata models share characteristics in their metadata elements, structures, and level of complexity for description. The design of the various metadata schemes has been largely influenced by USMARC. Although the various schemes are complementary, each also has its own emphasis.

For instance, USMARC records, TEI Headers, and EAD can provide highly structured bibliographic descriptions of electronic resources. However, USMARC also provides for summary bibliographic descriptions of print and electronic resources. TEI Headers are used to provide bibliographic and non-bibliographic description of SGML-conformant text, though it can also be used for non-SGML-conformant text. EAD allows for detailed description of archives and manuscript collections. It is also used for library finding aids, as it allows inventory description of the full range of archival holdings in various media.

Metadata Architecture

How can the above overlapping, complementary, or even competing metadata schemes communicate with one another regardless of location and syntaxes? Specifically speaking, how can those "aggregated metadata objects" [18] such as USMARC bibliographic records, SGML metadata records, Dublin Core metadata records, GILS records, finding aids in EAD, and other future metadata records be organized in a consistent way so that they can be interchanged in a distributed networking environment?

A metadata architecture is necessary to address the issues. Web Collections, [19] PICS (Platform for Internet Selection), [20] and Warwick Framework [21], and RDF (Resource Description Framework) [22] are examples of recent proposed metadata architecture derived from the need.

Metadata Repository

The metadata movement on the Internet points up the need for building a metadata repository. The respository should reside on the Internet. It should be extensive enough to incorporate many metadata formats into its databases and to be integratable with legacy systems such as library OPACs. The repository should be functional enough to facilitate portable metadata creation, sophisticated metadata searching, mapping and interchange, and easy updating and deletion.

It should also support the navigational and hierarchical architecture to 1) collocate different metadata types associated with an information object in diverse formats, and 2) maintain links from any of the metadata types to the object, as well as its related objects. Many types of information services should be supported at the interface level, for example information filtering, selecting, acquiring, organizing, delivering, billing, and so forth.

Then what should this metadata repository be, from the library's perspective? How can we ensure good access to the repository?

The Library Metadata Repository and the Library OPAC

Library OPACs have long been regarded as information gateways. The databases such as bibliographic databases, citation databases, and finding aids in which OPACs are directly indexed, or interfaced have served as library metadata repositories for years. So, can a library OPAC serve as a gateway to metadata repositories on the Internet as well as to its existing databases? The recent evolving technologies on the Web and the Internet, as well as in the computer industry have promised such potential.

The Library OPAC as Gateway to Metadata Repositories

The typical library OPAC gateway under development now by library communities is Window-interfaced with Z39.50 search capabilities, and TCP/IP connections. It provides users with a single interface to access library bibliographic databases, networked CD-ROM databases, full-text databases, online services, and other remote Z39.50-compatible databases. SilverPlatter's WinSPIRS [23] is an example of such a system.

The system has also been extended to Web-to-Z39.50 gateways. Web browsers become the clients of Z39.50 servers. As a result, OPAC databases are integrated with a wide variety of internal Internet resources, as well as external Internet resources. Yale University's WWW Orbits Gateway [24] is an example.

Evidently, with the demand for seamless access to diverse resources, using such a system to interface the metadata repository on the Internet seems feasible.

The Library Web Site as a Metadata Repository

A typical library Web site nowadays is still the place where a number of selected Internet resources, electronic journal collections, full-text databases, vendor-supplied Web-interfaced databases, and other library-related resources are mounted. [25]

Some libraries have a search engine built into their Web servers so that users can have full-text search features. The majority of them, however, still maintain lists of Web resources arranged by subject, or classification. The first approach doesn't permit targeted searching. In other words, it has low precision. The second approach requires users to be familiar with the collections, and to browse the listing page by page. For subjects with which the user is unfamiliar, access to the full extent of the subject is limited.

With the explosion of electronic information on the Internet, a library Web site becomes the only location for library electronic resources and services. With the capability of the Web-to-Z39.50 gateway, the Web site is broadened to legacy resources as well. The library Web site is no longer a mirror of a library, but an electronic library for diverse resources and services. A systematic and standard approach to all of the library's resources is required. The limitations of full-text searching, and of simply browsing a list of Web resources on a Web site, point out the advantages of using metadata.

In addition, with the increasing use of metadata, more and more Internet resources will have metadata already attached. Accordingly, libraries are obligated to maintain a metadata repository for Web resources. A library OPAC gateway will therefore become a logical system to access the repository.

Locally Created Metadata on Local Databases as a Metadata Repository

Many libraries are using metadata already. [26, 27, 28, 29] They have created EAD records for archival description and finding aids, SGML TEI Headers for electronic text, Dublin Core for simplified description of networked resources, and GILS for Government Information Locator Services. Most of these metadata can be created with local word-processing or database systems, but retrieval technologies are still under development, [30] and as a result, no common interface is in place yet for all of these metadata formats. For instance, GILS records are currently searched via Z39.50 gateways, Dublin Core Metadata Sets via automatic search engines, and EAD and SGML TEI Headers via SoftQuad's Panorama. Developing library OPAC gateways as a common interface to the database seems the only way to go.

Limitations of Web-to-Z39.50 Gateways

If library OPACs are used as a common gateway to all of the above metadata formats, Z39.50 profiles have to be established for each of them. Currently only MARC, GILS, and some vendor databases have such profiles.

Evidently, Web-to-Z39.50 gateways are not flexible enough to handle all necessary data types and structures. Therefore, if library OPACs are used as gateways to access all the databases, including metadata repositories either on library Web sites or on local databases, a metadata conversion system built into the gateway is needed to ensure metadata interchange.

Web-to-Z39.50 gateways are session-oriented. Even with Z39.50 profiles extended to other commonly accessed metadata formats, users will still have the problem of large system overhead, as the server remembers each session initiated, but never knows when to delete unnecessary sessions. With diverse metadata to be searched simultaneously, the slow response time could be a critical problem for effective retrieval.

Therefore, due to the limitation of current technology, using one type of metadata that has been profiled and implemented with Z39.50 as the base format for all the other metadata formats seems to be a practical approach. A metadata conversion system is then needed to convert all the other metadata to the base format, currently USMARC, in any future SGML-based encoding scheme.

For libraries which don't have Z39.50 capabilities, and who need to incorporate records encoded with other metadata schemes into their OPAC databases, a library metadata conversion system is needed as well.

In short, when a library OPAC is used as a gateway to access metadata repositories on its Web site or on local databases, a metadata conversion system is essential for three reasons:

- to capture, upgrade, and convert selected incoming metadata into the OPAC
- to integrate locally created metadata into the OPAC
- to display metadata in a user-specified format.

A Library Metadata Conversion System

Metadata Conversion

Metadata conversion for library resources means:

1) conversion of the selected incoming metadata into USMARC format;
2) conversion of locally created metadata into USMARC format, and;

3) conversion of metadata in metadata repositories into a user-specified format when records are being displayed.

Whether it is incoming or locally created, the metadata may be encoded in various formats. This paper deals only with recently proposed metadata standards for Internet resources, namely Dublin Core Metadata Sets, EAD, TEI Header, GILS, and the USMARC format. The conversion system will automatically remove the metadata from their heterogeneous encoding schemes, and convert them into a generic format—currently USMARC format—at the point of metadata integration.

The encoding level of both incoming and locally created metadata may differ as well. A lot of the literature has covered the metadata analysis aspects. [31, 32, 33] Too little meta-information will fail to meet the purpose of information discovery and retrieval. Too much meta-information is slow and costly to generate. Therefore, the essential metadata elements need to be defined for the conversion, and the conversion system must automatically verify these essential elements.

As most existing library resources have been encoded in USMARC format, a library metadata conversion system may still use the USMARC format as the basis for metadata description, storage, exchange, manipulation and retrieval. As a result, all metadata will be converted into USMARC format during integration and display. In this way, the existing bibliographic data can be preserved, and the full functionality of integrated library systems can be utilized.

Basic Functional Requirements

A library metadata conversion system must have the following basic functionalities (but not necessarily in this order):

<u>Extract Metadata</u>

The process is to capture metadata, and extract it when A) metadata is already associated with the information object either via hyperlinks or attached to the object, and B) highly structured information objects with no metadata attached.

In situation A, metadata can be extracted by automatically matching semantically similar elements and structures found in standard metadata format templates, namely the templates for Dublin Core Metadata Sets, EAD, TEI Header, GILS, and the USMARC format. If selected resources contain metadata that match with one of the templates, the metadata will be extracted and converted. If they do not match or just partially match with any of the templates, offline data mapping and modification are required. Li and Clifton's Semint is an example. [34]

Extracting metadata in situation B depends largely on the media and formats of the information object. Different media (such as text, images, and audio) and different formats (such as an HTML, SGML, PostScript, DVI, PDF, etc.) will require different programs to extract metadata automatically. Shklar et al.'s InfoHarness system is an example. [35] It is based on the fact that the lowest level of granularity of information is a logical unit of information, which they call an information unit (IU). The system includes an extraction program to recognize functional blocks (i.e. sections) within individual IUs, and then to encapsulate and index them.

Both Semint and InfoHarness have practical implications for automatic metadata extraction in library OPACs. Existing library databases are highly structured. They consist primarily of bibliographic databases with records in USMARC format. Selected Internet resources are usually in HTML or other structured file formats. Therefore, it is possible to extract metadata from the information object at the semantic level and information unit level, as in situations A and B respectively.

Map Metadata

This process finds corresponding attributes between extracted metadata and standard metadata templates. It occurs several times during the conversions:

- In the course of extracting, standard metadata templates are matched against metadata associated with selected Internet resources. The purpose is to identify content-bearing metadata elements, recognize metadata formats, and verify essential metadata attributes. Off-line adding and deleting data elements can also take place at this point.
- In the course of converting (see below), all extracted and verified metadata are mapped to a USMARC template, or a USMARC record is simulated into templates for other metadata formats. Each template contains essential data sets to describe information of various types, but the USMARC template serves as a common template for metadata comparison and exchange.

Convert Metadata

This process transforms metadata formats. A converter program takes the data elements from a newly extracted metadata record, and loads them into corresponding USMARC fields according to a specified USMARC template. Thus, a new USMARC record is created. The same method can be used to convert a USMARC record into other metadata formats.

Integrate Metadata

This process resolves the differences when combining data values and data structures that reflect the same information for the same entity from multiple databases. [36] It is a continuing process, involving extracting, mapping, converting, and indexing metadata into existing databases.

Provide an Interface to Metadata

This process provides an interface to facilitate a metadata search of both metadata repositories and existing library resources, and to display the search results in a user-specified metadata format. Using a library OPAC as a metadata interface will preserve the sophisticated search features and powerful linking models of the library OPAC, [37] and will also provide integrated searching of heterogeneous information in various metadata formats.

A Conversion Model for the Library OPAC

This model details the procedures of a metadata conversion system during the process of metadata extraction, mapping, conversion, and integration, in the three scenarios:

- For selected Internet resources which are to be integrated with existing OPAC databases, the incoming metadata on the repository (that is, the library Website) will be extracted in the following steps:

    1. It identifies the hyperlink associated to the page and takes information on the page as an

information unit (see the discussion of the IU above).
2. It maps semantically similar data elements and structures with standard metadata templates, e.g. Dublin Core, and expresses them as metadata.
3. It verifies the completeness of the data elements by mapping to its own template.
4. It loads the data elements into a USMARC template, converts them into a USMARC record, and integrates them into existing databases.
5. It searches the data elements using library OPACs.

- For locally created metadata on the repository (that is, specialized metadata mounted on local databases), the metadata conversion system will identify the content-bearing metadata elements, load them into a specified USMARC template, and convert and index them into existing databases;

- When a library OPAC is used as a gateway to access remote metadata repositories, a metadata conversion system will verify if the resources contain meta-information, load the data elements into metadata templates, confirm the metadata format and encoding level, and then display metadata in user-specified formats.

Benefits of the Model

The Internet contains various kinds of information in such diverse formats that it would take substantial *human* effort to create a metadata record, let alone to have it done automatically. This model therefore suggests building metadata extraction functions into the metadata conversion system, for three reasons:

- The model is intended for selected Internet resources which have metadata already attached. If these resources are selected as part of a library's collections, then those which have commonly used metadata attached could have that metadata captured, modified, converted, and integrated into the existing databases, either automatically or semi-automatically;
- The model is also designed for locally created metadata, such as finding aids in Dublin Core, EAD or SGML TEI Headers formats. These metadata are often generated with human intervention. They tend to have high-quality data contents, elements, and structures. Extracting, converting and integrating these metadata into the OPAC databases will ensure efficient metadata access;
- This model is applicable to Internet resources which lack metadata but which have highly standardized data attributes and structures. The system will identify a Web page and then extract meta-information from these attributes and structures, and thereby create a metadata record.

This model also suggests using the USMARC format as a common template for metadata comparison and exchange. One reason for this is that millions of existing library collections have already use USMARC as the metadata encoding standard. Converting other metadata standards to USMARC format allows integrated access to both the existing resources and newly selected resources.

Another reason for using USMARC as a common template is that the Electronic Location and Access (856) field has been added to USMARC, making it possible to connect USMARC records to their source data directly via sophisticated OPACs. In addition, the 856 field has been improved so as to accommodate Uniform Resource Names, to indicate file format type, and to indicate the relationship between the 856 and the resource which the record describes. [38] With a full Z39.50 connection, library OPACs becomes an ideal hub for information resources. Adding metadata conversion systems into the OPAC will further enhance interoperability among metadata repositories.

In addition, this model also handles metadata conversions in an integrated way. Other models deal with only one format at a time. This model can automatically detect which format to use by matching extracted metadata against a number of standard metadata templates simultaneously.

Furthermore, this model offers the library community the ability to choose a metadata standard that fits individual local needs. Catalogers can work with the metadata format which is easy for them to use, but later on have their records converted into another format which is compatible with their local databases.

Finally, this model will greatly increase the speed of adding filtered Internet resources into library bibliographic databases. Library users can obtain an accurate and complete picture of information on a given subject with a one-stop search on the library OPAC.

Issues for the Future

A number of issues still need to be addressed.

Semantic Differences

Whether extraction occurs at the data source level or metadata level, one of the biggest challenges for metadata extraction is semantic differences such as attribute naming, format, structure and identity conflicts. For example, attribute naming is problematic when data elements are conceptually the same but named differently, or named the same but conceptually different. [39] A 650 field in USMARC is named as LCSH, while in other metadata format it is named as keywords, key words, subject, subject headings, etc. Metadata naming standardization is a remedy.

Losing Metadata

During the metadata conversion, metadata elements in one format may not fit in another format. Vizine-Goetz, Godby and Bendig [40] and Xu [41] have done some studies on data element and data structure mapping among recently developed metadata standards such as the Dublin Core, TEI Header, and USMARC. They found that there were some overlap data sets among these standards, but not all of them had one-to-one correspondent matches. For instance, the TEI Header contains some data elements that are impossible to convert into USMARC. A global system control number is needed to keep track of the format change of each metadata record. This would enable the system to trace the USMARC record back to its original metadata format.

Web-Interfaced OPACs

Building a metadata conversion system at the Web-interfaced OPAC level is essential. Most current library metadata conversion programs are built on the Web, [42, 43] because that is where the resources reside (and should reside). A Web-interfaced OPAC will have the hyperlinking feature of the Web which connects the metadata and the information object. It also makes it possible for the conversion system to periodically and (semi) automatically verify source data against extracted metadata. [44] The process will ensure that the metadata still accurately reflects the information object it describes. Another reason is that many conversion programs are available in CGI (Common Gateway Interface) scripts. It is time saving to build the conversion systems upon existing programs. There are several projects involving automatic metadata conversion already underway. [45, 46]

Conversion and Indexing

With today's technology, the OPAC armed with a metadata conversion system will surely increase the speed of metadata integration and display. Will future technologies such as IIOP (Internet InterORB Protocol), [47] CORBA (Common Object Request Broker), [48] Java, and new Z39.50 profiles offer the same capability without conversion?

AACR and USMARC were originally designed for static information objects with flat data structure. However, with the impact of the metadata movement, will they be replaced by newly emerging metadata designed for network use, such as Dublin Core, [49] Warwick Framework, [50] PICS, [51] and XML (Extensible Markup Language)? [52] What impact would these formats have on conversion?

With a variety of metadata and metadata sets integrated into the OPAC databases, what will be the best indexing strategies for the OPAC to ensure effective access? Should there be one index for all media? Or multiple indexes for all various media? Or should both be allowed to exist?

Conclusion

By examining the metadata movement on the Internet, this paper points out the need to use the OPAC as the gateway to access to library metadata repositories. Within the context of these depositories, the paper analyzes a library metadata conversion system which will be able to extract selected incoming external metadata from Internet resources, convert it into USMARC format, and integrate it with existing library databases automatically. The model also suggests the possibility of transferring locally created metadata into the format that is the best to describe Internet resources, and then converting and integrating the resulting records into the OPAC databases. And finally, the metadata needs to be able to be displayed in a format which the user specifies.

With such a system in place, libraries would be able to expand their collections to incorporate a wide variety of information on the Internet with ease. They could share, re-use and integrate metadata repositories around the world.. And library users could have access to a wealth of diverse information resources.

These theories and procedures on how to build this system still need to be tested and proved. This paper has attempted to propose a model, and to highlight the importance of a metadata conversion system in the library.

---

1. "The Summary Object Interchange Format (SOIF)." Available: http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/user-manual/node151.html
2. L. Dempsey and R. M. Heery, "A Review of Metadata: A Survey of Current Resource Description Formats," in *DESIRE - RE 1004*, May 28, 1997. Available: http://www.ukoln.ac.uk/metadata/DESIRE/overview
3. "META Tagging for Search Engines." Available: http://WWW.Stars.com/Search/Meta/Tag.html
4. Willy Cromwell-Kessler and Ricky Erway, (meeting report from Metadata Summit organized by the Resarch Libraries Group, Mountian View, Calif., July 1, 1997).
5. Network Working Group, "Architecture of the WHOIS++ Service." Available: ftp://ds.internic.net/rfc/rfc1835.txt
6. "Lightweight Directory Access Protocol." Available: http://www.umich.edu/~rsug/ldap/ldap.html
7. "International Standard Z39.50: Information Retrieval (Z39.50): Application Service Definition and Protocol Specification." Available: http://lcweb.loc.gov/z3950/agency/

8. "SoftQuad's Panorama Publishing Suite." Available: http://www.sq.com/products/panorama/
9. "The TEI Header," in *TEI Guidelines for Electronic Text Encoding and Interchange*. Available: http://etext.virginia.edu/bin/tei-tocs?div=DIV1&id=HD
10. "Encoded Archival Description (EAD) DTD." Available: http://lcweb.loc.gov/ead/
11. "Harvard/Radcliffe Digital Finding Aids Project: History and Project Report." Available: http://hul.harvard.edu/dfap/projectdescription.html
12. David Beckett, "IAFA Templates in use as Internet Metadata." Available: http://www.w3.org/pub/Conferences/WWW4/Papers/52/
13. Stuart Weibel and Renato Iannella, "The 4th Dublin Core Metadata Workshop Report," in *D-Lib Magazine*, June 1997. Available: http://www.dlib.org/dlib/june97/metadata/06weibel.html
14. "Government Information Locator Service (GILS)." Available: http://www.usgs.gov/public/gils/
15. "Development of the Encoded Archival Description Document Type Definition." Available: http://www.loc.gov/ead/eadback.html
16. Michael Day, "Metadata: Mapping Between Metadata Formats." Available: http://www.uloln.ac.uk/metadata/interoperability/
17. Rebecca Guenther, "Metadata, Dublin Core and USMARC: a Review of Current Efforts." Available: gopher://marvel.loc.gov/00/.listarch/usmarc/dp99.doc
18. Carl Lagoze, Clifford Lynch, and Ron Daniel Jr., "The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata." Available: http://cs-tr.cs.cornell.edu:80/Dienst/Repository/2.0/Body/ncstrl.cornell%2TR96-1593/html
19. Yoelle Maarek and Dror Zernik, "Web Collections: A Mechanism for Grouping Web Documents and Site Mapping." Available: http://www-ee.technion.ac.il/W3C/WebCollection.html
20. "Platform for Internet Content Selection." Available: http:// www.w3.org/PICS/
21. Carl Lagoze, "The Warwick Framework: A Container Architecture for Diverse Sets of Metadata," in *D-Lib Magazine*, July/Aug. 1996. Available: http://www.dlib.org/dlib/july96/lagoze/07lagoze.html
22. Ralph Swick, "Resource Description Framework (RDF)." Available: http://www.w3.org/Metadata/RDF/
23. "Silverplatter World." Available: http://www.silverplatter.com/
24. "WWW Orbis Web-to-Z39.50 Gateway." Available: http://webpac.librry.yale.edu/webpac5.htm
25. "MIT Libraries." Available: http://libraries.mit.edu/
26. "Encoded Archival Description (EAD) Sites on the Web." Available: http://www.loc.gov/ead/eadsites.html
27. "Development of Integrated Online Community Information Systems." Available: http://www.solinet.net/monticello/monticel.htm
28. Ibid.
29. Andy Powell, "Metadata Management." Available: http://www.ukoln.ac.uk/metadata-june1997/ap/sld001.htm
30. "Z39.50 Profile for Access to Digital Library Objects: Fourth Draft." Available: http://lcweb.loc.gov/z3950/agency/profiles/dl.html
31. Rachel Heery, "Review of Metadata Formats," *Program* 30, issue no. 4 (Oct. 1996).
32. Eric Miller, "Monticello Electronic Library: Dublin Core Element Set Crosswalk." Available: http://www.oclcorg:5046/~miller/DC/crosswalk.html
33. "Digital Libraries: Metadata Resources." Available: http://www.nlc-bnc.ca/ifla/II/metadata.htm
34. Wen-Syan Li and Chris Clifton, "Semint: A System Prototype for Semantic Integration in Heterogeneous Databases," in *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, *May 23-25, 1995*, *San Jose*, *California*, (New York: Association for Computing Machinery, 1995).
35. Leon Shklar, Amit Sheth, Vipul Kashyap, and Kshitij Shah, "InfoHarness: Use of Automatically

Generated Metadata for Search and Retrieval if Heterogeneous Information," in *Advanced Information Systems Engineering: 7th International Conference, CAiSE '95, Jyvèaskylèa, Finland, June 12-16, 1995 : Proceedings* (New York: Springer, 1995).

36. Wen-Syan Li and Chris Clifton, "Dynamic Integration in Multidatabase Systems," *Journal of Database Management* 7, no. 1 (1996): 28-40.
37. Amanda Xu, "Hyperlinks in the WebPAC Systems: A Study of Hypertext Links to A Bibliographic Record in the Web Based OPACs," in *ISAL '96: Proceedings of the International Symposium on Academic Libraries in the 21st Century: Sept. 1-4 1996, Shanghai, China* (Shanghai: Shanghai Jiao Tong University Press, 1996).
38. See the *Guidelines for the Use of Field 856*, rev. Aug. 1997 (Washington, D.C.: Library of Congress, Network Development and MARC Standards Office). Available: http://www.loc.gov/marc/856guide.html
39. Stephanie Cammarata, Iris Kameny, Judy Lender, and Corrinne Replogle, "A Metadata Management System to Support Data Interoperability, Reuse and Sharing," *Journal of Database Management* 5, no. 2 (spring 1994): 30-40.
40. Diane Vizine-Goetz, Jean Godby, Mark Bendig, "Spectrum: a Web-Based Tool for Describing Electronic Resources," *Computer Networks and ISDN Systems* 27 (1995): 985-1001.
41. Amanda Xu, "Access Information on the Internet: Feasibility Study of USMARC Format and AACR2," in *Proceedings of the OCLC Internet Cataloging Colloquium, San Antonio, Texas, January 19, 1996*. Available: http://www.oclc.org/oclc/man/colloq/xu.htm
42. Robert Atkinson, "Meta Access Query," intercat@oclc.org (12 July 1996).
43. Gerald S. Barton, "Multiple Metadata Formats from the NOAA Environmental Services Data Directory," in *Proceedings of the First IEEE Metadata Conference, April 16-18, 1996, Silver Spring, Maryland* (New York: IEEE, 1996).
44. Pamela Drew and Jerry Ying, "A Metadata Architecture for Multi-System Interoperation," in *Proceedings of the First IEEE Metadata Conference*.
45. Stanley Elswick, Gary McCone, and Paul M. Kuin, "Report on Possible Conversion From the DIF to the USMARC Format," *Library Hi Tech* 13 no. 1-2 (1995): 36-42.
46. Jackie Shieh, "Common Good: Cataloging Operations and Electronic Text Processing," (paper presented at Finding Common Ground Conference, Harvard College Library, Cambridge, Massachusetts, Mar. 30, 1996).
47. "CORBA 2.0/IIOP Specification: the Complete Formal, 97-02-25: CORBA 2.0/IIOP Specification." Available: http://www.omg.org/corba/corbiiop.htm
48. David Curtis, "Java, RMI and CORBA." Available: http://www.omg.org/news/wpjava.htm
49. Committee on Cataloging: Description and Access, Task Force on Metadata and the Cataloging Rules, "Dublin Core and the Cataloging Rules: Analysis Project." Available: http://www.libraries.psu.edu/iasweb/personal/jca/dublin/index.htm
50. Carl Lagoze, "Rethinking Metadata." Available: http://www2.cs.cornell.edu/lagoze/talks/austalk/index.htm
51. Philip DesAutels, "PICS and the Dublin Core," (paper presented at Seminar on International Metadata Development, Mar. 6, 1997). Available: http://www.nla.gov.au/niac/meta/desautel/desautxt.html
52. "Specification for XML-Data." Available: http://www/microsoft.com/standards/xml/xmldata.htm