

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

LIS 571 Organization and Control of Recorded Information

Prof. D. Karpuk

4.25 / 5 = B

Interesting points
Do use standard spacing, & professionally present your paper
Examples warranted

Final Paper

4/30/01

Zapopan Muela

543 Baldy Hall Box 69

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

INTRODUCTION

Why the leading role of librarians in the metadata revolution?

Metadata revolution as we define it below is changing the librarians' paradigm of having the world of information under their control. But even with the new challenges and quests information represents to librarians, they have to keep on being the leaders in this information revolution. So this paper is here to emphasize and value the role of the librarians, particularly of the catalogers in their stake in the metadata revolution. Therefore, the reader will find in this paper a treatment more on the issue side than on technological discoveries.

I know Ben ←

Benjamin Hu, librarian, helps us to make our point more clear:

"Historically, librarians have organized the world's information. For centuries since the last information explosion (i.e. invention of printing), they have

developed the principles and bibliographic control that have proved successful in aiding searching, filtering, organizing and retrieving information of many types.”¹

ANALYSIS

What is metadata?

There are many definitions of metadata and the term has been in the library jargon for over a decade.

We have the feeling, as for the literature reviewed, that librarians have a general kind of fear to deal with the technological trends surrounding cataloging. They leave the technological developments to the computer scientists, that is, dealing with the issues of computers applied to the cataloging theory and practice. But that is or has to be—and that will be the guiding thread of this paper—a librarian’s fundamental function, specially a cataloger’s one.

For instance, we found several definitions of metadata on the computer science side and few on the librarian’s side. Taylor—a librarian—in her book: Wynar’s Introduction to Cataloging and Classification 9th edition, 2000, does not deal with metadata at all, and the term, theories and practices behind it, have been going on since quite a while ago. But she gives us a good definition in her glossary, and that may be in our analysis her only contribution to metadata:

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

"An encoded description of an information package (e.g. an AACR2 record encoded with MARC, a Dublin Core record, etc.); the purpose of metadata is to provide a level of data at which choices can be made as to which packages one wishes to view or search, without having to search massive amounts of irrelevant text" ² The metadata record for a digital object typically contains both descriptive and functional information. The item is described in a manner to facilitate discovery and assist collection maintenance. Metadata may include technical information about how an item may be used, as well as information about how an item may be used, and about the conditions of use too. ³

Metadata comes as an answer ^{to} the broader task of cataloging the Internet, the Internet resources, that is to control the ever in crescendo world of information being created, transmitted and in somehow organized on the Internet or particularly the World Wide Web or Web. Vinh-The Lam –a librarian– mentions that metadata systems and schemes have been created for this purpose, i.e., providing Internet resource creators with a working mechanism to embed resource description elements in their products. ⁴

Cataloging the Internet resources has been such a great challenge for librarians in the last years that The Haworth Information Press even created a new journal devoted for this task: Journal of Internet Cataloging. The International Quarterly of Digital Organization, Classification & Access.

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

If automating the catalogs long time ago with the creation of one of the first great metadata endeavors: the MACHine Readable Catalog record, back in the 1960s, was such a great challenge, to build also the first Online Public Access Catalogs (OPAC), then we can imagine that cataloging the Internet is even a greater challenge.

And what are some of the issues in cataloging the Internet that have taken a good deal of debate among librarians and non-librarians –specially from the computer science fields?

The librarians David Ward and Diane VanderPol illustrate us with some problems when cataloging Internet resources:

- The question of authority for the Web page the catalog entry refers to is resolved because the content creators and site maintainers are the library staff themselves.
- If the library Web page has moved for some reason, that decision can immediately be made known to the catalogers by the Web sites maintainers, and relevant OPAC changes made.
- The content-provider librarians would also be in charge of maintaining consistency on their individual pages, so that if a list of biology resources grew into a list of biology and botany resources, catalogers would be notified of the need for additional subject headings.

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

- Broken links discovered by link checking software run on the Web sites would necessitate only HTML level changes to the Web guide, not 856 field or other changes in the catalog entry itself. ⁵

We would like that those problems enlisted above would be the only ones around the great issue of cataloging the Internet or in other words the metadata revolution. David Ward and Diane VanderPol did not finish quite well their article when new technologies came to town. By the time they wrote their article a new technology called XML came aggressively as to overthrow HTML –the first generation of Web publishing, standing for HiperText Markup Language ⁶–. XML stands for extensible Markup Language. ⁷

The features of XML:

- While HTML was created for any computer user to read Internet documents, but faced the problems of incompatibility of computer systems. XML came to solve this problem, it makes more sense because consists of nothing but ordinary text.
- XML uses Unicode, ⁸ 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.
- One of the major characteristics for the library world is that allows another standards, like the Resource Description Framework (RDF) , to catalog Internet resources. ⁹

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

And with RDF we are entering the national baseball leagues. Tim Berners-Lee, non-librarian, the inventor of the Web and founder of the World Wide Web Consortium (W3W), ¹⁰ describes RDF as a: “scheme for defining information on the Web. RDF provides the technology for expressing the meaning of terms and concepts in a form that computers can readily process. RDF can use XML for its syntax and URIs ¹¹ to specify entities, concepts, properties and relations.” ¹² Norm Medeiros, librarian, adds up: “RDF serves as a structure into which any metadata semantic (such as Dublin Core) can operate. It supports semantic interoperability, that is, semantic elements can be “mixed and matched” within its framework while supporting the automated parsing of non-related schemes.” ¹³ And if this is not clear enough, Renato Iannella, non-librarian, adds in his “An Idiot’s Guide to the Resource Description Framework” article that “the consistent use of metadata and application of metadata schemas means that semantic interoperability will be preserved, hence significantly improving the deployability of advanced Web applications.” ¹⁴

Thus, we come to realize that cataloging the Internet has led us from HTML, through SGML, XML and RDF. And now we have a new technology called Semantic Web. But RDF in itself is not a semantic standard. Jean Hudgins, Grace Agnew and Elizabeth Brown, librarians, state in their book: *Getting Mileage out of Metadata. Applications for the Library, 1999* that RDF is just a metadata framework, but that can help in the creation of semantic metadata standards such as Dublin Core. They discuss that semantic metadata has to do

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

with semantic interoperability, “the transparent access to data can also be achieved at the semantic level though the use of a “least common denominator” metadata standard that defines core descriptive elements which can be extended across disparate subject and application domains.”¹⁵ They also highlight two standards for semantic interoperability: the Dublin Core and Field Mapping.

“The Dublin Core Metadata Element Set (DC)” –says Jian Qin, librarian— “is one of the metadata schemes used by the library and information community in creating metadata for networked information resources.”¹⁶ That is, like MARC worked and still works as metadata, data describing other data or surrogates, by means of its fixed fields, like author, title, subject, etc. of books, or serials, or videos, or CD-ROMs, or microfilms or other types of library materials on the shelves, the Dublin Core uses 15 elements resembling those of MARC to describe metadata or surrogates of a given record. From March 1995 to December 1996 the Ohio Computing Library Center (OCLC) developed their 15 elements of the Dublin Core Metadata Element Set: 1) title, 2) author or creator, 3) subject or keywords, 4) description, 5) publisher, 6) other contributors, 7) date, 8) resource type, 9) format, 10) resource identifier, 11) source, 12) language, 13) relation, 14) coverage and 15) rights. “The Dublin Core metadata element set is characterized by its simplicity, flexibility, and semantic interoperability, it provides the creators or publishers of digital information on the Web with an easy and established tool to create structured descriptions,” says Benjamin Ju, librarian.¹⁷

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

"Another method for providing semantic interoperability across metadata standard is "field mapping," or "crosswalks," which relate fields in one standard to their equivalents in another." ¹⁸ This field mapping or crosswalks deal for example when metadata systems like MARC are intended to migrate to a more sophisticated systems in an automated fashion as to preserve its original values and not to lose important information. ¹⁹

Dublin Core may sound to many an outstanding solution for cataloging the Internet, the Web. But the creators of the Astronomy Digital Image Library do not think the same. They describe "the efforts to bring scientific data into the digital library. This has required the extension of the standard WWW, and also the extension of metadata standards far beyond the Dublin Core. Our system demonstrates this technology for real scientific data from astronomy." ²⁰

By the time we are writing this —4/30/01—the most aggressive technology which has just arrived to this speed race is DAML. The DARPA Agent Mark Up Language developed by the US Defense Advanced Research Projects Agency. This is not the cutting but bleeding edge on semantic metadata, to find semantic terms is now not enough, but "to create technologies that will enable software agents to dynamically identify and understand information sources, and to provide interoperability between agents in a semantic manner." ²¹

CONCLUSION

What we have tried to show by little pieces is just the metadata revolution. The catalogers have a great challenge to try to play a leading role in this revolution.

These are the challenges:

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

- Make a solid input in the authority control over the Web. They must envision a new spectrum of authority control that includes many types of identifiers along with more familiar names, titles, and subjects.²²
- They have to watch for the unchangeable tasks of librarianship which are the development of subject headings and controlled vocabulary.

And like Thomas Mann says:

“The intellectual structure of cataloging is much more than any software, which determines the predictability, serendipity, and overall depth of a research library’s access systems to books.”²³

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

REFERENCES

- ¹ Hu, Benjamin. "Managing Pathways to Information Resources on the Web." *Journal of Educational Media and Library Sciences*. Vol. 37, No. 1, Sept. 1999, p. 1-16. Retrieved online 4/20/01 <http://vweb.hwwilsonweb.com>.
- ² Taylor, Arlene G. *Wynar's Introduction to Cataloging and Classification*. 9th ed. Englewood, Colorado: Libraries Unlimited, Inc., 2000, p. 501.
- ³ Cole, Tymothy W., Robert S. Allen and John G. Schmitz. "Building an Outreach Digital Library." *Illinois Libraries*. Vol 82, No. 4, Fall 2000, p. 241.
- ⁴ Lam, Vinh-The. "Cataloging Internet Resources: Why, What, How." *Cataloging & Classification Quarterly*. Binghamton, NY: The Haworth Information Press, Vol. 29, No. 3, 2000, 57.
- ⁵ Ward, David and Diane VanderPol. "Librarian, Catalog Thy Work! Getting Started Integrating Internet Resources into OPACs." *Journal of Internet Cataloging. The Internet Quarterly of Digital Organization Classification & Access*. Binghamton, NY: The Haworth Information Press, Vol. 3, No. 4, 2000, p. 58.
- ⁶ HTML: Hypertext Markup Language. The language used to encode formatting, links and other features on Web pages. Uses standardized "tags" such as <H1> and <BODY> whose meaning and interpretation is set universally by the World Wide Web Consortium. Definition taken from Berners-Lee, Tim, James Hendler and Ora Lassila. "The Semantic Web. A New Form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities." *Scientific American*. No. 0501, May 2000, retrieved online 4/23/01 <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.
- ⁷ XML: extensible Markup Language. A markup language like HTML that lets individuals define and use their own tags. XML has not built-in mechanism to convey the meaning of the user's new tags to other users. Definition taken from Berners-Lee, Tim, James Hendler and Ora Lassila. "The Semantic Web. A New Form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities." *Scientific American*. No. 0501, May 2000, retrieved online 4/23/01 <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- ⁸ Unicode: What is Unicode? Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language. Taken from: "What is Unicode" *The Unicode Standard*. Retrieved online 4/23/01 <http://www.unicode.org>.
- ⁹ Bosak, Jon and Tim Bray. "XML and the Second-Generation Web." *Scientific American*. No. 0599, 1999. Retrieved online 4/20/01 <http://www.scientificamerican.com/1999/0599issue/0599bosak.html>.
- ¹⁰ The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding. Retrieved online 4/23/01 <http://www.w3.org/>.
- ¹¹ URI: Universal Resource Identifier. URLs are the most familiar type of URI. A URI defines or specifies an entity, not necessarily by naming its location on the Web. Definition taken from Berners-Lee, Tim, James Hendler and Ora Lassila. "The Semantic Web. A New Form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities." *Scientific American*. No. 0501, May 2000, retrieved online 4/23/01 <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>
- ¹² Idem.
- ¹³ Medeiros, Norm. "XML and the Resource Description Framework: the Great Web Hope." *Online*. Weston, Conn: Vol. 24, No. 5, S/O 2000, 37-40. Retrieved online 4/20/01 <http://vweb.hwwilsonweb.com>.

THE LEADING ROLE OF LIBRARIANS IN THE METADATA REVOLUTION

By: Zapopan Martin Muela-Meza

¹⁴ Iannella, Renato. "An Idiot's Guide to the Resource Description Framework." *The New Review of Information Networking*. Vol. 4, 1998. Retrieved online 4/20/01 <http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/>.

¹⁵ Hudgins, Jean, Grace Agnew and Elizabeth Brown "Interoperability. Creating a Shared Information Environment for the 21st Century" in *Getting Mileage out of Metadata: Applications for the Library*. Chicago: American Library Association, 1999 (LITA Guides No. 5), p. 36.

¹⁶ Qin, Jian. "Computational Representation of Web Objects in an Interdisciplinary Digital Library: A Survey and Experiment in Polymer Science." *Annual Review of OCLC Research* 1998. Retrieved online 4/20/01 <http://vweb.hwwilsonweb.com>.

¹⁷ Hu, Benjamin. "Managing Pathways to Information Resources on the Web." *Journal of Educational Media and Library Sciences*. Vol. 37, No. 1, Sept. 1999, p. 1-16. Retrieved online 4/20/01 <http://vweb.hwwilsonweb.com>.

¹⁸ Hudgins, Jean, Grace Agnew and Elizabeth Brown "Interoperability. Creating a Shared Information Environment for the 21st Century" in *Getting Mileage out of Metadata: Applications for the Library*. Chicago: American Library Association, 1999 (LITA Guides No. 5), p. 36.

¹⁹ McCray, Alexa T. and Marie E. Gallagher. "Principles for Digital Library Development." *Communications of ACM*. Vol. 44, No. 5, May 2001, p 50- 51.

²⁰ McGrath, Robert E., Joe Futrelle, Ray Plante and Damien Guillaume. "Digital library technology for locating and Accessing Scientific Data." *Proceedings of the Fourth ACM Conference on Digital Libraries*, 1999, p. 188.

²¹ Handler, James. "DARPA Agent Mark Up Language (DAML)." *ISO World Programs*. Retrieved online 4/23/01 <http://www.darpa.mil>.

²² Vellucci, Sherry L. "Metadata and Authority Control." *Library Resources and Technical Services*. Vol. 44, No. 1, Jan 2000. Retrieved online 4/20/01 <http://vweb.hwwilsonweb.com>.

²³ Mann, Thomas. *Library Research Models: A Guide to Classification, Cataloging, and Computers*. New York, Oxford: Oxford University Press, 1993, p. 147.