Introduction

To metadata or not to metadata, that is the question.
Whether 'tis nobler in the mind to suffer the slings and arrows of outrageous search results
Or take up metadata against a sea of irrelevance
And by organizing them, find them?[1]

I had to include this quote; it was so original and so apropos for this report. It's from Tom Reamy's October 2004 piece in *EContent*, "To Metadata or Not to Metadata."

Reamy's article comes from the corporate environment, a landscape facing problems not dissimilar from the ones the library community faces when it comes to metadata. Reamy's article on metadata is one of the best discussions—outside of librarianship—I have encountered thus far on this subject. In his article, Reamy deals with the issue of cost, and he discusses the pros and cons of doing it well—as opposed to doing it poorly. He also mentions how the corporate community has started to question the value of adding metadata at all.

For Reamy, there are three approaches for the business/corporate environment when it comes to implementing and incorporating metadata:

- hire consultants (high upfront costs and ongoing maintenance costs);
- have content publishers create their own metadata (low quality, low participation); or
- automatic metadata generation (high costs and unreliable).

Given the problems with these three options, Reamy suggests another approach, which was brought up at the 2003 Dublin Core Initiative workshop: the content-value-tier model, first discussed by well-known information architecture expert Lou Rosenfeld. This model goes something like this: Don't try to solve all the world's problems, just focus on practical solutions and high-value content (using criteria such as currency, reusability, authority, strategic value, and popularity). In other words, for high-value content, use full metadata; for low-value content, use little or no metadata. Yet even this approach has its problems. Who decides what is high- vs. low-value content? And this solution doesn't even begin to deal with the political and territorial issues rampant in most businesses (sounds familiar, doesn't it?).

"To Metadata or Not to Metadata," EContent Magazine*, October 2004*
www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=7118

Reamy divides the practical solution into three dimensions: intellectual infrastructure, information infrastructure, and infrastructure context. Issues such as people, keywords, and other minefields are examined.

In the end, it seems corporations just can't commit to huge investments—such as a corporate taxonomy, an enterprise-side metadata standard, implementation retroactively on thousands of documents—and then the integration of *that implementation* on *current* projects and practices. So, from Reamy's and Rosenfield's experiences, the best result toward incorporating metadata into the corporate environment goes something like this:

1. Create an overall infrastructure vision, including metadata standards.

2. Create a team of people, with official recognition in their job descriptions and with a reward structure, to implement that vision.
3. Regarding the metadata itself, buy/customize/mine an existing taxonomy or controlled vocabulary, rather than create your own.

To conclude, Reamy summarizes these steps as: Think Big, Start Small, Scale Fast. His final quote deserves mention, as you don't often hear this type of statement in the corporate community:

> You wouldn't think of running a company without organizing your employees, why do you think you can create access to information without organizing that information?[2]

There are lessons to be learned from Reamy's article. The most important: There's an entire world *outside* of librarianship trying to tackle the issue of information organization and description. An article appearing on *IEEE Distributed Systems Online*, "Web Metadata Standards: Observations and Prescriptions," is an example of this.[3]

In this summary article of current Web metadata standards, standards such as ebXML (Electronic Business using eXtensible Markup Language), WSDL (Web Services Description Language), UDDI (Universal Description, Discovery, and Integration), and P3P (Platform for Privacy Preferences) are listed with the more familiar DC (Dublin Core), SOAP (Simple Object Access Protocol), and OWL (Web Ontology Language). The author prescribes lessons from software engineering, software reuse and library science, and artificial intelligence in tackling the issues related to choosing a single metadata standard when serving multiple purposes. His prescriptions are valid and viable in the library environment as well, although librarians rarely venture outside of their field when researching information on what they feel are their areas of expertise.

"Is It Time for a Moratorium on Metadata?" is a wonderful article by Dick C.A. Bulterman in the "Visions and Views" section of *IEEE Multimedia* (October-December 2004) regarding the viability of metadata at all.[4] He begins with a wonderful parable concerning breadmaking, geography, popularity, and marketing; you have to read it to understand its applicability to metadata in the twenty-first century.

Although Bulterman's fairly simplified description of the metadata environment in the last ten years can be argued, his overall assessment of the usefulness of metadata should be considered.

I like his personal definition of metadata, "optional structured descriptions that are publicly available to explicitly assist in locating objects." His sections, "Creating Metadata for Text Has Gone from Tedious to Insignificant"; "Creating Metadata Description Is an Error-Prone Task"; and "Creating Metadata: Context-Sensitive, Culturally Biased, and Time-Variant," lead the reader to his final pronouncement, ". . . to save metadata, we first need to ignore it."

"Is It Time for a Moratorium on Metadata?"
IEEE Multimedia, Oct-Dec 2004
http://homepages.cwi.nl/~dcab/PDF/ieeeMM2004.pdf

"Meta Meta Data Data: Making a List of Data about Metadata and Exploring Information Cataloging Tools," by Ralph Kimball
www.fortunecity.com/skyscraper/oracle/699/orahtml/dbmsmag/9803d05.html

"Web Metadata Standards: Observations and Prescriptions," IEEE Software*, Jan/Feb 2005*
http://dsonline.computer.org/portal/site/dsonline/menuitem.6dd2a408dbe4a94be487e0606bcd45f3/index.jsp?&pName=dso_level1_article&TheCat=1005&path=dsonline/0502&file=s1bod.xml&

Bulterman then issues a five-point plan for a moratorium on metadata, simple in its proclamation and impossible in its implementation. What is interesting, however, is that point number four ("Ask public-spirited citizens worldwide to contribute their favorite photos, audio fragments, or personal videos to create a culturally diverse corpus of 1 million nontext media assets. . . .") is currently happening without any restraints or restrictions (see the discussion on folksonomies, p. 49). The author's realignment with the breadmaking parable from the beginning of the article brings this rather revolutionary work to its conclusion.

Ralph Kimball, well-known data warehouse architect, provides an informative discussion on metadata in the data warehouse community in, "Meta Meta Data Data," an article that appeared in the March 1998 issue of *DBMS Magazine* (now known as *Intelligent Enterprise*, www.intelligententerprise.com).[5] Although it's dated, the article does provide some clues into what's happening outside our field in the area of metadata.

The purpose of beginning this introduction with these non-library articles on metadata is to illustrate the fact that there's an entire world outside of librarianship dealing with the same issues and problems that we are. Are we working with them? Are we forming collaborative partnerships or consortia to work through and perhaps experiment with alternative and viable solutions to the Google dilemma? (That is, the notion that Google's and other search engines' proprietary algorithms and searching mechanisms are "good

enough" for everyone.) After all, most of the world has sent a notice to librarianship: "We are happy getting thousands of hits to our information queries," and "We are happy to be able to get instant gratification and instant access to resources that may not be exactly what we are looking for but are 'good enough' (and especially fast enough) to find answers to and resources about our questions/problems/research."

It is the business and corporate sectors that have the money and the means for experimentation and problem solving; they also have the technical and marketing experience to get the job done. So why aren't we working with them to develop more viable and better quality information organization and retrieval methods for the Internet? Probably for the same reasons that caused much of the current metadata standards development outside of librarianship in the late twentieth century: We either ignore them, or we don't actively work with them because we are too pigeonholed in our traditional structures and methods.

Since the publication of this title in 2002, much more definition and specialization regarding the concept of metadata has taken place. Now there's general agreement there are four layers of standards that support metadata: data structure or schema standards (examples include Dublin Core and VRA Core); data communication or encoding standards (such as MARC and XML); data content standards (like AACR2 and CCO); and data value standards (LCSH, AAT are examples here). Data structure and data communication standards can be combined into a single standard (like EAD or TEI), and data communication and data content standards are often used together to create a metadata standard (such as MARC and AACR2). Discussions on metadata creation now focus on who and how: Who will do it (professionals, technical creators like Web masters and encoders, content creators, community or subject enthusiasts, or a combination of or all of the above)?; and How will it be done (templates, editors, markup tools, extraction, conversion, generators, or a combination of or all of the above)?

More metadata standards have been formalized, using the International Standards Organization/International Electrotechnical Commission (ISO/IEC) 11179 standard, "Metadata Registries Standard" (accessible at http://metadata-stds.org/11179).

This standard includes extensive instructions for how to identify data elements and register a scheme with a registration authority. The standard is divided into six parts: 1) Framework, 2) Classification, 3) Registry Metamodel and Basic Attributes, 4) Formulation of Data Definitions, 5) Naming and Identification Principles, and 6) Registration. The ISO/IEC 11179 has become essential to the database community, and today is a vital resource for the development of metadata schemes for digital resources.

This report will provide new directions and updates to the 2002 report; both reports should be consulted together as a somewhat total picture and research guide on metadata.

In this report, Chapter 1, "General Resources," provides an update on important general resources currently available on metadata (without rehashing old material in the 2002 report).

ISO/IEC 11179 "Metadata Registries (MDR) Standard"
http://metadata-stds.org/11179

Chapter 2, "Update on Major Metadata Standards," examines the metadata standards discussed in the 2002 report and provides information on major new resources and news currently happening with these standards.

Chapter 3, "New and Emerging Metadata Standards," examines new and emerging metadata standards not mentioned in the 2002 report and follows the same structure of presentation and information as that report. (At the end of Chapter 3, there's s a list of miscellaneous resources that didn't seem to fit in anywhere else.)

Chapter 4, "Current Issues and Developments Related to Metadata: Thoughts and Opinions" is a presentation of my thoughts and opinions of current issues and developments related to metadata as well as important resources for more information. These topics are: metadata management; the rise of personalized information organization systems called "folksonomies"; automatic metadata generation; controlled vocabularies and ontologies; weblogs; metadata harvesting; and 2D/3D information visualization applications.

Finally, Chapter 5, "Metadata Futures," lists some resources that discuss metadata futures.

Notes

1. Tom Reamy, "To Metadata or Not to Metadata," *EContent Magazine* (October 2004), www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=7118&IssueID=211 (accessed September 27, 2005).
2. Ibid.
3. Mordechai Ben-Menachem, "Web Metadata Standards: Observations and Prescriptions," *IEEE Software* 22, no. 1 (January/February 2005), http://dsonline.computer.org/portal/site/dsonline/menuitem.6dd2a408dbe4a94be487e0606bcd45f3/index.jsp?&pName=dso_level1_article&TheCat=1005&path=dsonline/0502&file=s1bod.xml& (accessed September 27, 2005).
4. Dick C.A. Bulterman, "Is It Time for a Moratorium on Metadata?" *IEEE Multimedia* (October/December 2004), http://homepages.cwi.nl/~dcab/PDF/ieeeMM2004.pdf (accessed September 27, 2005).
5. Ralph Kimball, "Meta Meta Data Data," *DBMS Magazine* (now *Intelligent Enterprise*, www.intelligententerprise.com), March 1998, www.fortunecity.com/skyscraper/oracle/699/orahtml/dbmsmag/9803d05.html (accessed September 27, 2005).