Chapter 4

Current Issues and Developments Related to Metadata

Thoughts and Opinions

In this chapter some of the current developments related to metadata will be examined. Many of them are much broader than just standards development; some of these issues are metadata management, the rise of personalized information organization systems known as "folksonomies," automatic metadata generation, controlled vocabularies and ontologies, weblogs, metadata harvesting, and 2D/3D information visualization applications.

Metadata Management

Three years ago, I predicted that the number of metadata standards would decrease rather than increase as time progressed. Although many similar standards in subject disciplines have been merged and combined, there are still many metadata standards out there today. General agreement has been reached regarding Dublin Core as the lowest-common-denominator metadata standard available, useful for large-scale description and harvesting of information regarding digital objects, especially through the OAI-PMH protocol. But other standards continue to be refined and revised for the long haul (EAD, TEI, MARC-based XML standards, ONIX, LOMS, RDF, etc.), while other types of metadata (administrative, structural, technical, etc.) continue to proliferate and be incorporated both within and outside of library-based information organizations.

The issue of the management of metadata, both in terms of volume (number of records) and in terms of interoperability (number of standards), is becoming a major priority for larger academic institutions and organizations. Coupled with this are issues related to searching and retrieving large data warehouses and repositories of harvested metadata and how to do this as efficiently as possible. Variability of user interaction and uses of these repositories for different functions and research needs also dictate that the design of specialized and unique retrieval interfaces may well be a future job duty for librarians.

Martin Kurth and his colleagues at Cornell University have had to deal with this issue more so than most organizations. Because the OAI-PMH protocol and its creators and designers are located at Cornell, and much of the experimentation and harvesting using OAI is being done there, they are dealing with many of the issues related to metadata management well ahead of everyone else. In the article "Repurposing MARC Metadata: Using Digital Project Experience to Development a Metadata Management Design" (*Library Hi Tech* 22:2, 2004, p. 153–65), Kurth and his colleagues examine current challenges to metadata management at Cornell.[1] The issues include workflows not well established, the fact that mapping and transformation of metadata is decentralized, and that documents and files are located all over the place.

The Cornell experience illustrates how complex and large metadata operations will have to deal with both number of records and many different types of metadata standards in the future. Kurth and colleagues demonstrate how the XML/XSLT approach assists them in metadata management, as well as how important it is to coordinate metadata creation/manipulation and to manage the tools and files that result from metadata processes. In designing a metadata management strategy for the future, Kurth indicates the following:

- inventory metadata relationships and processes organization-wide;
- identify data stewards for mapping and transformations resources;
- promote sharing and reuse of metadata resources to reduce costs and risk; and

- use documented relationships and processes to coordinate metadata activities.

For Cornell, the next steps are to continue to build library consensus about metadata element sets and mappings; to meet with stakeholders to discuss the creation of a MARC repurposing design structure; to investigate the creation of a repository of metadata schemes, tools, and files; to develop reusable transformation tools; and to extend stakeholder discussions to a library-wide metadata management design. Kurth's PowerPoint presentation related to this issue (which is based on Kurth's and his colleagues' *Library Hi Tech* article) given at the ALA 2005 Annual Conference in Chicago can be found at www.ala.org/ala/alcts/alctsconted/alctsceevents/alctsannual/MODS_Kurth.pdf.[2]

OCLC has also been researching and examining the issue of metadata management in the future. OCLC's Research Division has published "Metadata Switch: Thinking about Some Metadata Management and Knowledge Organization Issues in the Changing Research and Learning Landscape" (accessible at www.oclc.org/research/publications/archive/2004/dempsey-mslitaguide.pdf). This assessment of the current and future environment surrounding metadata management reveals there will be many contact points with digital objects from the management point of view, which will ultimately result in: multiple metadata creation and repository environments and multiple metadata formats in these multiple environments; the need for multiple controlled vocabularies to organize and access the content; the growing need and interest in automatic categorization and creation of metadata; and developed methods that instruct how to manage complex digital objects. In thinking about these issues, other concerns arise, such as dealing with provenance and trust of shared metadata, managing identity and difference, consistency and cost, forms of disclosure, the harvesting of content and complex objects, the many-to-many issue, and gated resources and authorization. These are only a few of the concerns addressed in this article; there are short-term solutions provided, and long-term goals and objectives are discussed.

As the volume of metadata and the number of harvested repositories increase, and with the explosion of information resources that will continue to appear and grow in networked and online environments, metadata management will become more of a challenge and focus. Librarians need to be at the forefront of examining solutions and directions in this area.

Weblogs

Weblogs, known commonly as "blogs" and "blogging," have risen to prominence and importance in this early part of the twenty-first century. In a world where information is power, keeping track of current news, breaking stories, and professional literature is vital. Blogging has become one of the new communication modes in recent years, and technology has provided this innovation by allowing one to receive this information on his or her desktop instantaneously–from the moment of publication by a blogger.

Metadata plays a key role in the organization and delivery of blog information (see Metadata for Weblogs in Chapter 3). Without metadata, it would be difficult for the massive amounts of bytes and briefs continuously posted on blogs to be sent to subscribers or to be found in archives through Google and other search engines.

For a history of Weblogs and blogging, as well as a long list of types of blogs, see the article in Wikipedia at http://en.wikipedia.org/wiki/Blog.

Metadata Management Resources

"Using MARC Repurposing to Initiate a Metadata Management Design," by Martin Kurth
www.ala.org/ala/alcts/alctsconted/alctsceevents/alctsannual/MODS_Kurth.pdf

I would not have been able to write this report without the help of blogs. In this fast-paced Web technology environment, where it is imperative for us to stay current with the professional literature, blogs have immediately notified me when items appear on the Internet, when they are published, and provided me with immediate access to important items–all without having to scour through professional literature, greatly diminishing the probability that I will only find out about a pertinent article months after it is published.

This has been especially helpful in the world of metadata standards, where news about these standards, or some new information, is published/released almost every day.

I use Bloglines to manage my blog subscriptions (a free service available at www.bloglines.com; there are many others out there). As I locate or hear about blogs in my areas of interest, I can add them to my list of blogs. When a new post appears, I am notified through Bloglines, and my icon in the lower right portion of my computer screen indicates that I have new postings. I then click on the icon, which opens a separate window containing all of my blog subscriptions, with the blogs highlighted that contain new messages and information. See the chart on pages 47–48 for a list of current blogs that I maintain and receive information through.

Related to blogs are wikis. In education and collaborative learning, wikis also assist with cooperative

learning and interactive communication. There are many resources available to find out about wikis; a quick guide available from EDUCAUSE, "7 Things You Should Know about Wikis," is available at www.educause.edu/ir/library/pdf/ELI7004.pdf.

OpenURL Metadata Sharing

Through the application of OpenURL-based metadata sharing, one new initiative is attempting to bridge the gap that exists among all of the new and hot technology tools available right now (like blogging and link logging) with library services. There are two distinct components of the OpenURL framework; the first is the address of an OpenURL resolver. The second component is the bibliographic metadata that describes the relevant article. In "Opening Up OpenURLs with Autodiscovery" (accessible at www.ariadne.ac.uk/issue43/chudnov), the authors examine, through experimentation and example, how OpenURLs can create new information pathways for libraries to incorporate for their users.[4] In their conclusion, the authors indicate there are "too many standards and service models finding too few users." In fact, here is an extensive quote from this article that all libraries should take notice of:

> . . . we believe a new type of information resource optimization is now needed to help fix the problem of insufficient integration. It is no longer sufficient to work only to improve human interfaces for usability and wayfinding, and separately, to improve automated interfaces for harvesting and scalability. We now also need to design our services to accommodate the needs of both users and systems to move freely between human and automated interfaces . . . we now need to design for the likelihood that users enter our resources from a wide variety of directions: from search engines, from their friends' weblogs, from course pages, from sharing networks, and from hand-coded Web pages. . . . Above all, we must find ways to help them do all of that with only one or two clicks. Lorcan Dempsey at OCLC has referred to this technique of better

Blog Title	Blog Content	Blog URL
Beyond the Job	professional development blog for librarians	http://librarycareers.blogspot.com
Catalogablog	library cataloging, classification, metadata, etc.	http://catalogablog.blogspot.com
Chronicle.com—The Wired Campus	education and tech news from *The Chronicle of Higher Education*	http://wiredcampus.chronicle.com
Chronicle.com—Today's News	current news from *The Chronicle of Higher Education*	http://chronicle.com/news
DAVA—Digital Audiovisual Archiving	digital transformation & preservation of audiovisual material	http://av-archive.blogspot.com
Darcusblog	geek tools and the scholar	http://netapps.muohio.edu/blogs/darcusb/darcusb
The Digital Library Sphere	projects, standards, research, etc.	http://digitallibrarysphere.blogspot.com
Digitization 101	issues related to digitization	http://hurstassociates.blogspot.com
Diglet	UCSD Digital Library Planning Working Group	http://gort.ucsd.edu/mtdocs/diglet
Enterprise Metadata	R. Todd Stephens's blog on metadata	www.rtodd.com/blog

Blog Title	Blog Content	Blog URL
The Google Weblog	new stuff on or about Google	http://google.blogspace.com
Hangingtogether.org	information from RLG (Research Library Group)	www.hangingtogether.org
Infomusings Blog	musings on library and information science	www.infomuse.net/blog
inSilico	Princeton University Library metadata blog	http://libserv12.princeton.edu/insilico
The Kept-Up Academic Librarian	news and developments in higher education	http://keptup.typepad.com/academic/
Librarian's Index to the Internet	new and newly discovered Web sites for librarians	http://lii.org
Library Grants	for librarians interested in grant opportunities	http://librarygrants.blogspot.com
Library Journal: Digital Libraries	information related to digital libraries	www.libraryjournal.com/community/891/Digital+Libraries/42816.html
Library Journal: InfoTech	information related to information technology	www.libraryjournal.com/community/891/InfoTech/42857.html
Library Stuff	resources for professional development	www.librarystuff.net
LISNews.com	librarian and information science news	www.lisnews.com
Loomware: Crafting New Libraries	discussions related to information and technology in academic libraries, universities, and life in general	http://blog.uwinnipeg.ca/loomware
News from CNI	(Coalition for Networked Information)	www.cni.org
Outgoing	library metadata techniques and trends by Thom Hickey at OCLC	http://outgoing.typepad.com/outgoing
Peter Scott's Library Blog	Peter is best known as the compiler of the first hypertext index of Internet resources, Hytelnet, which was first released in 1991.	http://blog.xrefer.com
ResourceShelf	resources & news for information professionals	www.resourceshelf.com
Reuters: Top News	news from Reuters	http://today.reuters.com/news/default.aspx
Theoretical Librarian	Gerry McKiernan's blog	http://theoretical-librarian.blogspot.com

> preparing resources and services to be sewn together in new ways as 'intrastructure' (rather than 'infrastructure').[5]

Putting all of this together using metadata illustrates the power that structured organization and description of resources provides in the networked environment. Further building on the idea that OpenURLs can provide a Web-based mechanism to package and transport citation metadata so that users in libraries can more easily access the cited works, a specification has been written to imbed OpenURL metadata in HTML. Known as ContentObject in Span, or COinS, it is currently in version 2.0 and available at http://ocoins.info.

Folksonomies/Personalized Information Organization

In a world of user-directed information searching and user-constructed interfaces for better retrieval of information, it was only a matter of time before the users took things into their own hands. When it comes to organizing their information, information organizations have always *imposed their organization systems* for books, videos, CDs, digital objects, and everything else. Enter personalized information organization systems, instituted by the user, for the user, and for other users. These "classification/metadata" systems are also known as folksonomies, social bookmarks, ethnoclassification, and tagging. From the users' points of view, they are in charge of choosing appropriate terms for their digital objects, which are then shared in a community of users, who have the option of contributing their own terms for their own digital objects. Eventually, a huge metadata vocabulary builds up, which the users hone and change, depending on their own initiatives.

For instance, if someone has pictures of his or her dog online and decides to call the pictures "Dog," and someone else shares pictures of his or her dog online, and decides to call them "Sparky," is there a problem? Well, it depends on what the user wants. If the "Sparky" person wants people to know that his or her pictures are of his or her dog, and if the "Dog" person eventually discovers that people are searching for dogs under the term "dogs," both of these users have the option of changing their tags to the popular term in use—or not.

In this way, the metadata vocabulary ebbs and flows daily, constantly in flux and never static, always in tune with what users feel is the appropriate terminology. Of course, there are many drawbacks: no pre-determined standardization, no management, and no set organization. But then, this development is truly user-initiated, is driven by the users, and is ultimately *used* by the users.

And it is this last part that really makes the difference as to whether this "fad" will ultimately be successful or not. Remember when search engines appeared, and libraries didn't think that users would settle for thousands of hits and very little precision? Remember when Google began to be seriously used by millions of library users, and libraries didn't think that users would settle for "just good enough" when it came to searching the Web? Well, we have been proved wrong so far. It behooves the profession to constantly monitor the folksonomy world, because "the user always knows best."

Why do personal information organization systems work? Because they are responsive to change and context-specific; they are user-defined and they reflect the vocabulary of authors/users; they have the ability to uncover marginalized fields of knowledge; they use feedback to improve and manage the system; and they don't privilege one hierarchy or worldview over another.

What are some of the problems with folksonomies? Tags are often ambiguous, synonyms and plurals cause problems, scalability is difficult, and they produce a different type of chaos than pre-coordinated metadata.

Where are the users going to participate in tagging? There are a number of major Web sites devoted to sharing digital objects on the Web, where the users can supply their own tags, or metadata, at the time that they link their files. Some of the more popular sites are detailed below.

del.icio.us, accessible at http://del.icio.us, is a social bookmarks manager. It's a free service that allows users to add Web pages to their personal links, to tag those Web sites with their own keywords, and then to share their collection(s) with other users and other machines. On the right-hand side of the Web page are the most active tagging Web sites.

Flickr, accessible at www.flickr.com, is a shared Web space for photographs and digital objects on the Web. It is a free service; one just needs to set up a personal account and log in. To see how Flickr works, just click on the "Explore" button, then click on the "Most Recent Uploads" button. This will take you to the most recent photographs added, where you can click on "Popular Tags." Here you can see what a folksonomy really looks like. It is fluid, ever changing, and never static. The box shows the popularity of the tag through its size, as compared to other tags. These tags are clickable, so that users can access all photographs that use that tag as a metadata element.

CiteULike, accessible at www.citeulike.org, is a free online service that enables users to share, store, and organize academic papers. On the right-hand side of the Web page, the most active tags are shown, and their popularity is again indicated by their size in comparison to other tags. This is an example of how citation metadata allows for linking in the online environment and how the OpenURL initiatives (presented earlier in this report)

Resources on Folksonomies

del.icio.us
http://del.icio.us

Flickr
www.flickr.com

CiteULike
www.citeulike.org

Technorati
www.technorati.com

Furl
www.furl.net

Unalog
http://unalog.com

"Metadata for the Masses," by Peter Merholz
www.adaptivepath.com/publications/essays/archives/000361.php

"Folksonomies: Cooperative Classification and Communication through Shared Metadata," by Adam Mathes
www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

D-Lib Magazine *April 2005*
www.dlib.org/dlib/april05/04contents.html

"Taxonomies and Tags: From Trees to Piles of Leave," by David Weinberger
www.hyperorg.com/blogger/misc/taxonomies_and_tags.html

"Folksonomies? How about Metadata Ecologies?", by Louis Rosenfeld
www.louisrosenfeld.com/home/bloug_archive/000330.html

"Explaining and Showing Broad and Narrow Folksonomies," by Thomas Vander Wal
www.vanderwal.net/random/entrysel.php?blog=1635

"Folksonomies to Organize the News," by Monique Van Dusseldorp
www.poynter.org/column.asp?id=31&aid=79865

"In the New Game of Tag, All of Us Are It," by Stephen Levy
http://msnbc.msn.com/id/7445653/site/newsweek

"Social Bookmarks So Far," on Library Clips
http://libraryclips.blogsome.com/2005/04/15/social-bookmarks-so-far

can make this technology available to library users more directly within the library.

To see the power of **Technorati,** accessible at www.technorati.com, click on the "Tags" button in the top menu bar, and you will see an alphabetical list of current tags being used on the Web site. Again, size of the word indicates the popularity of use.

Frassle, accessible at http://frassle.rura.org, is an open source tool that provides an integrative blog environment, in which users can publish both links and original text in their own blogs. There are a number of features available, including an RSS aggregator. This social bookmarking Web site appears to be better organized on the front Web page than most others.

Information about **Furl** is available at www.furl.net/index.jsp. Furl stands for "file URL," and it was designed to save Web pages and to provide full-text searching for previously read materials. There is a very good guide to Furl at www.classroomhelp.com/workshop/Furl_Guide.pdf.

Unalog, http://unalog.com, is a social bookmarking tool written in Python and available as open source software. The primary aim is to make bookmarks available to other users on the system. It has the ability to slice and dice further queries, and supports private entries, private users, and public or private groups.

There have been many articles on this new phenomenon in the Web environment. What I consider to be the ones of major value are:

Peter Merholz, "Metadata for the Masses," *Adaptive Path* Essay Archives, October 19, 2004, www.adaptivepath.com/publications/essays/archives/000361.php (accessed October 6, 2005).

Merholz provides a simple explanation and discussion of ethnoclassification, using del.icio.us and Flickr as examples. One of the benefits of free tagging that he examines is the idea of "desire lines," trails that demonstrate how users choose to move as opposed to how they are told to move. This is shown graphically in everyday life by pathways that users build in public parks through normal use, as opposed to the paved paths built by the city. Merholz provides a graphic photo example of this in real life.

Adam Mathes, "Folksonomies: Cooperative Classification and Communication through Shared Metadata," Paper for Computer Mediated Communication (LIS590CMC), Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, December 2004, www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed October 6, 2005).

Adam Mathes's paper is a very convincing and detailed examination of the whole concept behind folksonomies and shared metadata. The author wrote this paper for a graduate class in computer-mediated communication. It divides metadata into three areas: professional creation, author creation, and user creation. It is the third area that is the topic of the paper. Mathes examines many of the pros and cons already considered regarding tagging, but he also mentions some unanticipated uses, as well as areas for further research. It is well written and scholarly in its approach on this topic.

The **April 2005 issue of *D-Lib Magazine*** (www.dlib.org/dlib/april05/04contents.html) contains an editorial and two excellent articles on social bookmarking tools. The first bookmarking article (accessible at www.dlib.org/dlib/april05/hammond/04hammond.html) provides a general overview of tagging, indicating it's a much more flexible and cheaper approach to classification than the top-down, traditional approach.[6] The authors provide some very good examples and charts illustrating how tagging works on the Web, their differences and similarities in relation to content creator and tag user, audience, and how they appear on the Web. They do not see these folksonomies as replacements for formal classification; rather, they are a supplemental means to organize information and order search results. In describing these new types of link managers on the Web, the authors provide a list of elements that distinguish folksonomy-based Web sites. The list includes:

- personal user accounts;
- browser bookmarklets to facilitate entry;
- searching by tag or user;
- RSS feeds;
- extensions and plug-ins;
- querying of links based on popularity, users, tags, etc.;
- classification by tagging; and
- mechanisms for entering links, titles, and descriptions.

The authors also provide a detailed list of current bookmarking Web sites and tools, and they even review and compare the features of some of the better-known ones.

In Part II (accessible at www.dlib.org/dlib/april05/lund/04lund.html), the authors focus on one specific service—Connotea—as a free online service for scientists.[7] Added benefits of this service have appeared as more individuals use it, such as tag convergence, recommendations, and directory creation.

As Bonita Wilson states in the opening editorial (accessible at www.dlib.org/dlib/april05/04editorial.html): "Individuals who are not information professionals are organizing and categorizing large amounts of external information both for their own use and for use by others.... Will it work?"[8] The end result remains to be seen.

David Weinberger, "Taxonomies and Tags: From Trees to Piles of Leaves," Introductory Section in *Release 1.0*, ed. Esther Dyson (February 2005), www.hyperorg.com/blogger/misc/taxonomies_and_tags.html (accessed October 6, 2005).

In David Weinberger's excellent article that relates tags to trees and leaves in the real world, knowledge becomes a forest, where it is now autumn and the leaves are falling, because traditional knowledge hierarchies do not work in the digital environment.

The author indicates there are three intellectual orders: the first, when objects were physical and we put them on shelves; the second, when we physically separated the metadata from the objects and created catalogs; and the third, when both the object itself and the metadata are digital. The author goes on to explore the history of tagging, going back to the concept of bookmarking in early Web browsers. The differences between folksonomies (self-organizing taxonomies) and faceted systems and tree structures (hierarchical taxonomies) are explained. In summary, the author discusses the pros and cons of these three approaches to organizing information.

Louis Rosenfeld, "Folksonomies? How about Metadata Ecologies?" LouisRosenfeld.com Blog, (January 6, 2005), www.louisrosenfeld.com/home/bloug_archive/000330.html (accessed October 6, 2005).

This post on the LouisRosenfeld.com blog is an interesting example of how communication and discussion develops and works within a weblog. Rosenfeld is well known as an information architect, and his comments and opinions regarding folksonomies and tagging are informative and interesting. He indicates that both traditional and newer forms of metadata-controlled vocabularies are needed in the digital environment. Neither work well alone, and together, they help to balance the pros and cons of each approach. Comments that follow this discussion provide insight and varying opinions and also show how this new technology helps to form community and informative debate in real time. One of the bloggers mentions an interesting metadata generation tool named "Meta-Door" and provides a link to information about it.

Thomas Vander Wal, "Explaining and Showing Broad and Narrow Folksonomies," vanderwal.net (February 21, 2005), www.vanderwal.net/random/entrysel.php?blog=1635 (accessed October 6, 2005).

Thomas Vander Wal's blog posting provides graphic illustrations and definitions of broad and narrow folksonomies on the Web. He explains that del.icio.us is a broad folksonomy approach, while Flickr is a narrow folksonomy approach.

"Tagging Helps Unclutter Data," CNN.com (May 3, 2005), no longer available on CNN.com.

This CNN Internet news story, dated May 3, 2005, is no longer available on the CNN Web site, but a search engine query did generate this URL, http://mailman.uba.ar/pipermail/infoedu/2005-May/000407.html (accessed on October 6, 2005), which, at the time of writing, featured the article's content. The article is another good explanation of tagging, current tagging Web sites, and why people are interested in doing it.

Monique Van Dusseldorp, "Folksonomies to Organize the News," Poytneronline.com (March 16, 2005), www.poynter.org/column.asp?id=31&aid=79865 (accessed October 6, 2005).

This article is a blog posting by Monique Van Dusseldorp on the Poytneronline.com weblog ("a group weblog from the sharpest minds in online media/journalism/publishing") regarding the current interest in tagging. It features some good links to various Web sites and articles.

Stephen Levy, "In the New Game of Tag, All of Us Are It," *Newsweek* (April 18, 2005), http://msnbc.msn.com/id/7445653/site/newsweek (accessed October 6, 2005).

In his "The Technologist" column on MSNBC.com and in the *Newsweek* Science and Technology section, Stephen Levy (using library jargon and references to the Dewey Decimal Classification System) explains "freestyle labeling" (the author refers to it as "tagging") and explains the differences between the two approaches (freestyle as opposed to the Dewey Decimal Classification approach).

"Social Bookmarks So Far," Library Clips Weblog (April 15, 2005), http://libraryclips.blogsome.com/2005/04/15/social-bookmarks-so-far (accessed October 6, 2005).

This blog post from the Library Clips weblog provides an excellent collection of online resources on the topic of social bookmarks. The various sections, with links to references, include: folksonomy vocabulary, extensible system, and innovative ideas. Although it was impossible to find the blog author's full name (listed as John T., a.k.a., johnt, in Technorati), this is good starting point for anyone beginning to research the topic of folksonomies.

Ontologies and Controlled Vocabularies

In relation to the discussion on folksonomies, the topic of ontologies and controlled vocabularies, as important contributors to the success of both tagging and metadata on the Web, has also arisen. Although folksonomies and ontologies are topics that support and help define the other, resources and discussions on ontologies (in relation to metadata) are becoming the focus of many metadata standards; they are being examined in relation to the current fads related to free tagging and personalized information organization initiatives. As such, it is important to understand that creating new ontologies, and supporting access to already constructed ontologies and controlled vocabularies, are now becoming the focus of many groups. Resources include:

Nick Mote, "The New School of Ontologies," paper by Nick Mote, www.isi.edu/~mote/papers/Folksonomy.html (accessed October 6, 2005).

Nick Mote provides an excellent introduction to the concepts behind the old and new schools of onotologies in his online paper, "The New School of Ontologies." The Web page features good quality references and links as well.

The **Living Taxonomy Project** (LTP), accessible at http://livingtaxonomy.org/index.php/Main_Page, is an open source, collaborative effort aimed at creating a global set of standards-based taxonomies for education. The idea currently is to construct a free (in the sense of open and collaborative, not in the monetary sense, although that is also implied) cataloging structure for sharing and collecting education materials worldwide. There are currently nine taxonomies posted for review and editing, and these are changing daily. There is an associated blog and RSS feed for those who wish to participate or stay current on this project.

The Technical Advisory Service for Images (TASI), "**Control Your Language—Links to Metadata Vocabularies**," accessible at www.tasi.ac.uk/resources/vocabs.html, provides a well-constructed portal to known ontologies and controlled vocabularies available on the Web. The sections are divided into thesauri, subject headings, and word lists; classifications (general and specialist); and authority lists (name and place).

The **OCLC Terminologies Pilot Project**, accessible at www.oclc.org/productworks/terminologiespilot.htm, is a recently announced pilot project to provide access

to multiple thesauri to help information organizations create consistent metadata for their collections. More information for getting access to this pilot project is available at the Web URL listed above.

Resources on Ontologies/Controlled Vocabularies

"The New School of Ontologies," by Nick Mote
www.isi.edu/~mote/papers/Folksonomy.html

The Living Taxonomy Project (LTP)
http://livingtaxonomy.org/index.php/Main_Page

"Control Your Language—Links to Metadata Vocabularies"
www.tasi.ac.uk/resources/vocabs.html

OCLC Terminologies Pilot Project
www.oclc.org/productworks/terminologiespilot.htm

Controlled Vocabularies and Folksonomies: Why Change Is Good," by Joshua Porter
http://bokardo.com/archives/change_is_good

"Published Subjects: Introduction and Basic Requirements"
www.ontopia.net/tmp/pubsubj-gentle-intro.htm

Metadata for the Alexandria Digital Library (ADL) Feature Type Thesaurus
www.alexandria.ucsb.edu/gazetteer/FeatureTypes/FTT_metadata.htm

"Here Is a How to Topic Maps, Sir!", by Alexander Johannesen
www.shelter.nu/art-007.html

"Metadata and the Web," by Mehdi Safari
www.webology.ir/2004/v1n2/a7.html

Joshua Porter, "Controlled Vocabularies and Folksonomies: Why Change Is Good," bokardo, a blog about Web 2.0 (January 28, 2005), http://bokardo.com/archives/change_is_good (accessed October 6, 2005).

This blog post provides discussion on the topic of how tagging is changing the landscape of information organization and why this is good. The problems of: single placement of an object, adding new content, unanticipated change in language and vocabulary usage, and embracing change are all examined. Associated postings regarding this essay are attached.

Steve Pepper, ed., "Published Subjects: Introduction and Basic Requirements," OASIS Published Subjects Technical Committee Recommendation (June 24, 2003), www.ontopia.net/tmp/pubsubj-gentle-intro.htm (accessed October 6, 2005).

OASIS is one of many established metadata standards currently working on an approach to controlled vocabularies for use within the OASIS standards. The OASIS Published Subjects Technical Committee has developed lists for countries and languages.

Metadata for the Alexandria Digital Library (ADL) Feature Type Thesaurus, Alexandria Digital Library, University of California at Santa Barbara, www.alexandria.ucsb.edu/gazetteer/FeatureTypes/FTT_metadata.htm (accessed October 6, 2005).

This is an example of one digital project's development of a thesaurus to assist the creators in their incorporation and use of metadata. Many other digital projects are beginning to do this as well.

Alexander Johannesen, "Here Is a How to Topic Maps, Sir!" www.shelter.nu (undated post), www.shelter.nu/art-007.html (accessed October 6, 2005).

Topic maps are related to ontologies and controlled vocabularies, in that they are graphical representations of these systems in practice. Alexander Johannesen's blog posting/essay examines topic maps in relation to the subject of music, and some of the difficulties and challenges that this particular subject discipline imposes on topic map representation.

Mehdi Safari, "Metadata and the Web," *Webology* 1, no. 2 (December 2004), www.webology.ir/2004/v1n2/a7.html (accessed October 6, 2005).

This article is listed in Chapter 1 of this report as an excellent essay about how metadata is the future of the Web. Linked to this author's opinion is the concept of how ontologies are the new form of semantic metadata for the new form of the Web. The bibliography section supports the author's assumptions and provides good further reading on the topic.

Automatic/Semiautomatic Metadata Generation

Perhaps the biggest issue in metadata standards today is the concept of automatic or semiautomatic processes for metadata generation. As is well known, metadata creation by professionals is expensive. For the last thirty years, library-cataloging operations worldwide have taken advantage of outsourcing their redundant and sometimes original cataloging of materials in an effort to

reduce costs, reduce backlogs, reduce staff, and increase efficiency. As more materials are acquired that are digital in origin and dissemination, the same problems transfer into the digital arena. In order to provide access to digital objects and resources, the quality of the metadata becomes highly important.

But who is supposed to create this metadata? If professionals do it (i.e., library catalogers), the cost becomes enormous. If the authors of digital objects do it, what about the quality of the metadata? What if they don't do it? If metadata can be generated automatically, what about quality and variability of the technology involved? These are some of the questions that have taken center stage in regard to digital object access (and their associated metadata for access and preservation) for the future.

There are many current initiatives, within librarianship and outside of it, focusing on the automatic or semiautomatic generation of metadata. Most libraries have experimented with cataloging selected online resources within their OPACs or in setting up aggregator lists or dynamic databases that provide access to their digital resources. But as Web and digital resources grow, along with unique digitization and digital collections, these methods will soon reach their limits.

Much of the research is being done on two fronts: within library science programs and/or among library science faculty and graduate students, and within computer science programs and/or among computer science faculty and graduate students.

The library science approach focuses more on the quality of the metadata and the tools to generate the metadata, as well as the various approaches involved with automatic and semiautomatic generation (i.e., human intervention and participation in the process).

The computer science approach focuses on the algorithms and finding the right mathematical formulas to build the machines to automatically generate the metadata. Both fronts, of course, tread on the other's areas of expertise, and work together when possible to address the challenges and technologies. To put the approaches into a more concise framework, library science addresses the people and metadata quality issues; computer science addresses the design and construction of the machines, programs, and tools.

The goal of the **AMeGA (Automatic Metadata Generations Applications) Project** (accessible at http://ils.unc.edu/mrc/amega.html) is to identify and recommend functionalities for applications supporting automatic metadata generation in the library and bibliographic control community. Based at the University of North Carolina, School of Information and Library Science, and led by Dr. Jane Greenberg, AMeGA is examined in a ninety-seven-page final report (www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf).[9]

Automatic/Semiautomatic Metadata Generation Resources

AMeGA Project
http://ils.unc.edu/mrc/amega.html

AMeGA Final Report
www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf

"Automatic Metadata Generation & Evaluation," by Elizabeth Liddy et al.
http://delivery.acm.org/10.1145/570000/564464/p401-liddy.pdf?key1=564464&key2=0115994211&coll=GUIDE&dl=GUIDE&CFID=53189359&CFTOKEN=39598383

"Iterative Design of Metadata Creation Tools for Resource Authors," by Jane Greenberg et al.
http://dc2003.ischool.washington.edu/Archive-03/03greenberg.pdf

"Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications," by Jane Greenberg
http://ils.unc.edu/mrc/automatic.pdf

"Usability of a Metadata Creation Application for Resource Authors," by Abe Crystal and Jane Greenberg
http://ils.unc.edu/~acrystal/crystal_greenberg_2005_LISR.pdf

There is a Metadata Generation Task Force (MGTF) that assists in this effort, with the project being conducted in connection with Section 4.2 of the Library of Congress Bibliographic Control Action Plan. The project goals (taken directly from the Web site) are:

1. To evaluate current automatic metadata generation applications. (The categories of tools being investigated include: document presentation software, tools created specifically for metadata generation, and online library cataloging modules for creating metadata.)
2. To survey metadata professionals to determine which aspects of metadata generation are most amenable to automation and semiautomation. Other metadata creators may also participate in the study.
3. To compile a final report of recommended functionalities for automatic metadata generation applications. The final report will be reviewed and endorsed by the Metadata Generation Task Force (MGTF) and be made publicly accessible via the Library of Congress.

(The project submitted its final report to the Library of Congress in February 2005.)

In terms of Project Goal #1, a literature review in the area of automatic metadata generation was conducted; functionalities of seven different types of content creation software were reviewed; a comparison of two current automatic metadata generation application tools, DC-dot and Klarity, was composed; a list of automatic metadata generation features advertised by selected ILS vendors was compiled; and an interview with a cataloger who had used one of the more advanced ILS systems was conducted.

For Project Goal #2, a survey was developed to gather data identifying desirable system functionalities for automatic metadata generation applications. This was restricted to digital document-like objects (DDLOs), defined as a primarily textual resource accessible through a Web browser. The summary of the survey indicates that most participants clearly supported automatic metadata generation as an executable first step in metadata creation, but then, a means should be provided for human evaluation and manual intervention as a second step.

For Project Goal #3, a final report was produced (www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf) with recommendations. The document is divided up into the following summary of findings: System Goals, General System Recommendations, System Configuration, Metadata Identification/Gathering, Support for Human Metadata Generation, Metadata Enhancements/Refinement and Publishing, Metadata Evaluation, and Metadata Generation for Nontextual Resources. The report ends with three main tasks for the Library of Congress to implement: build an automatic metadata generation application, foster and facilitate research on automatic metadata generation, and implement mechanisms for communicating and negotiating with content creation software vendors.

This document should be required reading for anyone involved in metadata creation, maintenance, or standards development. It is full of detailed information regarding the current state of development in automatic metadata generation applications, in both the library science and computer science fields. The information and analysis provided regarding the survey participants, and the services and tools they need in this area, is vital for any company or person designing these software applications. The bibliography is especially thorough.

The Web site (http://ils.unc.edu/mrc/amega.html) for the MRC at UNC-SLIS lists a number of metadata projects, in addition to the AMeGA Project. These include Metadata Generation Research (MGR), which is a broader research project in automatic and semiautomatic metadata generation that has funding support from Microsoft and OCLC; an Ontologies and Semantic Web Research group that is examining challenges within this topic; the OpenKey Project, with its major goal to create two polycave plant keys and produce an information repository of high-quality botanical images; and the User-Centered Metadata Research Project, which specifically looks at constructing retrieval systems for metadata geared towards users' information needs.

Hui Han et al., "Automatic Document Metadata Extraction using Support Vector Machines," Proceedings, 2003 Joint Conference on Digital Libraries, May 27–31, 2003, Rice University, Houston, Texas.

This article typifies research and development currently happening in the computer science field in the area of automatic metadata generation applications, especially for digital libraries. The authors identify several methods for automatic extraction of metadata: regular expressions, rule-based parsers, and machine-learning techniques (such as Support Vector Machines, statistical methods, symbolic learning, Hidden Markov models, inductive logic programming, and grammar induction). This paper examines Support Vector Machines, and then it goes into very detailed computer science vocabulary to describe this process.

Elizabeth D. Liddy et al., "Automatic Metadata Generation & Evaluation," Proceedings, 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland (August 11–15, 2002), http://delivery.acm.org/10.1145/570000/564464/p401-liddy.pdf?key1=564464&key2=0115994211&coll=GUIDE&dl=GUIDE&CFID=53189359&CFTOKEN=39598383 (accessed October 6, 2005).

This poster session abstract describes a project that is examining the automatic indexing of electronic resources and producing metadata records that enable users to retrieve relevant resources. It describes the testing and results of the project's research.

Jane Greenberg et al., "Iterative Design of Metadata Creation Tools for Resource Authors," Unattributed paper available (at the time of writing) at http://dc2003.ischool.washington.edu/Archive-03/03greenberg.pdf (accessed October 6, 2005).

I'm not quite sure whether this was published in a journal or not, as the PDF file is widely available on the Web without any publisher attribution. In the paper, Greenberg and her colleagues describe in detail the metadata-creation tools needed by resource authors. She divides them into three types: templates (simple forms), editors (forms enhanced with documentation and sometimes automatic processes), and generators (automatic or semiautomatic applications that require the submission of a URL). Examples are provided in the appendices of templates constructed for the National Institute of Environmental Health Sciences (NIEHS) project that funded this work.

Jane Greenberg, "Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications," *Journal of Internet Cataloging*, v. 6, no. 4 (2003): 59–82, http://ils.unc.edu/mrc/automatic.pdf (accessed October 6, 2005).

This article details the study done for the AMeGA project comparing DC-dot and Klarity, in terms of their capabilities as automatic metadata-generation applications. Greenberg provides several ideas on how these applications can be improved and suggests future areas of research. She concludes by stating that integrating extraction-harvesting methods appears to be the best approach to creating optimal metadata.

Abe Crystal and Jane Greenberg, "Usability of a Metadata Creation Application for Resource Authors," *Library and Information Science Research*, v. 27, no. 2 (2005): 177–89, http://ils.unc.edu/~acrystal/crystal_greenberg_2005_LISR.pdf (accessed October 6, 2005).

This article is another take on the NIEHS project described earlier.

2D/3D Information Visualization

The human race is one that is tremendously influenced by sensory perceptions. The way we as human beings understand, learn, grow, and adapt is based on our ability to perceive, view, and conceptualize thoughts and ideas. Even the terms we use to describe the process of learning and understanding revolves around the word "see." When an idea pops into our heads, or when a new concept moves from static jargon to understandable knowledge, we "see" it. Ideas become "clear," concepts are brought into "focus," and "a picture is worth a thousand words." The use of visual metaphors to describe cognitive relationships and human thought processes are interwoven into our daily lives and activities.

Recent advances in the sciences, technology, and commerce have accelerated the human desire to concentrate and focus on visual representation and metaphor in daily life. The availability and powerful capacity of computers, the explosion of information access and graphics via the Internet and media conglomerates, and the technological visual interfaces now becoming available through high-definition and plasma screens only fuel our desire and, indeed, our addiction for more visual and sensory input from many different sources.

Perhaps the major difficulty we now face is an increasing amount of nonlinearity and complexity in our lives from this technology, which produces a world of counterintuitive inputs and outputs. The power to visualize and graphically represent results, ideas, solutions, and problems—not just as a flat one-dimensional presentation, but in two or multiple dimensions—as well as design and present collaborative dimensional spaces in which more than one individual can contribute to problem-solving and idea-generation thought processes, is now a distinct possibility in the very near future. It is in this environment that the field of information visualization can be found.

Information visualization as a field is relatively new. Its foundational period is now ending, and it is rapidly moving forward into the marketplace. Since beginning in the 1980s—with high-end, expensive computer workstations that allowed for real-time and advanced interactive graphics for animation, space exploration, and visual effects in two-dimensional (hereafter 2D) and three-dimensional (hereafter 3D) formats—information visualization technologies are readily available to anyone with a standard PC platform. In fact, there are numerous newly launched software companies that focus specifically on the mass marketing of information visualization products, services, and experiments. There are many who believe that information visualization is poised to go from its current anonymity (in medical and scientific applications) to enter the mainstream of application design and user interface for anyone with access to a PC.

Within this focus on the visual and the graphic, metadata will be playing an essential if not primary role. A very dynamic and interesting article on current developments in this area is by Bernard Frischer.[10]

2D/3D Information Visualization Resources

"The Ultimate Internet Cafe: Reflections of a Practicing Digital Humanist about Designing a Future for the Research Library in the Digital Age," by Bernard Frischer
www.clir.org/pubs/reports/pub129/pub129.pdf

"From CVR to CVRO: The Past, Present, and Future of Cultural Virtual Reality," by Bernard Frischer et al.
www.cvrlab.org/research/images/CVR%20to%20CVRO.pdf

In his section on current activities at the Cultural Virtual Reality Laboratory (CVRLab) at UCLA, Frischer makes these statements about metadata:

> We must publish not only the 3-D data about an archaeological site but also the footnotes, or metadata, that tell users everything they might like to know about the reconstruction, from who made it to why one kind of marble or plant material was used instead of another. By publishing the metadata along with 3-D data, the CVRLab wants to enable users to distinguish the securely

known from the hypothetically reconstructed, to be aware of current scholarly controversies, and even to empower users to tear apart a model and put it back together in a way that seems more cogent. In developing metadata standards, the CVRLab and similar laboratories around the world are taking advantage of groundwork laid by librarians for the Dublin Core Metadata Initiative and the Visual Resources Association. On its own metadata committees, the CVRLab is seeking the active participation of librarians and information scientists.[11]

Frischer consistently mentions the importance of metadata in many of his publications regarding virtual reality and its future in libraries and on the Internet. In discussing the evolution of cultural virtual reality (CVR) into a professional organization that he names the Cultural Virtual Reality Organization (CVRO), Frischer details extensively the importance of metadata in our future virtual reality worlds:

> . . . Thus, the philologist editing a text deals with metadata in the preface, in the *apparatus criticus*, and in the text itself through the use of signs. . . . Accuracy pertains to the data and metadata; authenticity to the user's experience of the data and metadata. Research on other media suggests that, ironically, far from getting in the way of the user's experience and sense of authenticity, metadata can even add to the credibility of a CVR model, as can be seen in the following study of Condry (1989) cited by Shapiro and McDonald (1995, 338). . . .[12]

Given these statements, it would appear that metadata—its role in futuristic initiatives and endeavors related to electronic and virtual environments—is very important. It behooves us, as librarians, to continually monitor and actively participate and share our information-organization knowledge and experience with the broader world and the current marketplace. If we don't, others—who don't necessarily know or understand these issues but who have persuasive marketing skills and talents that place them at a better advantage—will. In the end, we will be the ones cleaning up the messes and failures, while others garner accolades and attention for the successes that metadata can bring to future information landscapes.

Notes

1. Martin Kurth, "Repurposing MARC Metadata: Using Digital Project Experience to Development a Metadata Management Design," *Library Hi Tech* 22, no. 2 (2004): 153–65.
2. Ibid., "Using MARC Repurposing to Initiate a Metadata Management Design," ALA Annual Conference 2005, PowerPoint presentation (June 27, 2005), www.ala.org/ala/alcts/alctsconted/alctsceevents/alctsannual/MODS_Kurth.pdf (accessed October 5, 2005).
3. Lorcan Dempsey et al., "Metadata Switch: Thinking about Some Metadata Management and Knowledge Organization Issues in the Changing Research and Learning Landscape," *EScholarship: A LITA Guide,* ed. Debra Shapiro, (Chicago: American Library Association, 2005), www.oclc.org/research/publications/archive/2004/dempsey-mslitaguide.pdf (accessed October 6, 2005).
4. Daniel Chudnov et al., "Opening Up OpenURLs with Autodiscovery," *ARIADNE* 43 (April 2005), www.ariadne.ac.uk/issue43/chudnov (accessed October 6, 2005).
5. Ibid.
6. Tony Hammond, Timo Hannay, Ben Lund, and Joanna Scott, "Social Bookmarking Tools (I): A General Review," *D-Lib Magazine* 11, no. 4 (April 2005), www.dlib.org/dlib/april05/hammond/04hammond.html (accessed October 6, 2005).
7. Ben Lund, Tony Hammond, Martin Flack, and Timo Hannay, "Social Bookmarking Tools (II): A Case Study—Connotea," *D-Lib Magazine* 11, no. 4 (April 2005), www.dlib.org/dlib/april05/lund/04lund.html (accessed October 6, 2005).
8. Bonita Wilson, "Personalized Information Organization," *D-Lib Magazine* 11, no. 4 (April 2005), www.dlib.org/dlib/april05/04editorial.html (accessed October 6, 2005).
9. Jane Greenberg, Kristina Spurgin, and Abe Crystal, "Final Report for the AMeGA Project," School of Information and Library Science, University of North Carolina at Chapel Hill (submitted to Library of Congress on February 17, 2005), www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf (accessed October 6, 2005).
10. Bernard Frischer, "The Ultimate Internet Cafe: Reflections of a Practicing Digital Humanist about Designing a Future for the Research Library in the Digital Age," *Library as a Place: Rethinking Roles, Rethinking Space,* Council on Library and Information Resources, Washington, DC (February 2005): 41–81, www.clir.org/pubs/reports/pub129/pub129.pdf (accessed October 6, 2005).
11. Ibid., p. 47.
12. Bernard Frischer et al., "From CVR to CVRO: The Past, Present, and Future of Cultural Virtual Reality," Quote material: 7–8, www.cvrlab.org/research/images/CVR%20to%20CVRO.pdf (accessed October 6, 2005).