Metadata: The Magic Behind the Scenes

By Sheila Bair, Metadata Cataloger & Pam Rebarcak, Head, Technical Services

For centuries, librarians have described and organized physical containers of information—scrolls, papyri, cuneiform tablets, manuscripts, non-print media, and, of course, the printed book. Monastic librarians arranged volumes simply by size—big books on the bottom shelf and small ones above—or even by color. Later, numeric and alpha-numeric schemata married the hefty tome to the slender pamphlet. Content dictated juxtaposition on any given shelf, which thereby became, visually, that bar code of varicolored spines that modern library patrons take for granted. But what happens when content escapes its container? In the digital library, this very eventuality has come to pass. The pages are, as it were, out of the binding or box. Only the content matters. Metadata makes it possible to find, link, and share information on a granular level regardless of its format.

Notice that we’ve said “metadata makes” and “metadata is.” Although the purist writer or editor may still consider data and agenda to be plural nouns, many 21st century authors consider metadata to be a singular noun. So, what is metadata? The Greek prefix tells the story: it means beyond, about, sometimes behind—as in metaphysics, metempsychosis, and so on.

Data Behind Data

Metadata is not, however, randomly accumulated data, but rather, according to Priscilla Caplan, writing in Metadata Fundamentals for All Librarians (ALA: 2003), structured information recorded with a documented metadata scheme. Metadata is information that describes or identifies other information (or another information source). Metadata is ordered according to one of several standard schemes. Such metadata or information about information becomes increasingly valuable in a knowledge economy. Metadata is the data working quietly behind information that makes it work and enables people and systems to do smarter things. The faster and more efficiently you can get the information you want or need, the less effort you waste—and the smarter and richer you become.

Still perplexed? The National Science Digital Library Metadata Primer defines metadata as “structured, standardized descriptions of resources, whether digital or physical, that aid in the discovery, retrieval and use of those resources” [http://metamanagement.com/nsdl/lib.org/overview2.html#what].

Exploding by the nanosecond, information threatens to overwhelm the frail thread of our capacity to order it. But information seekers are rarely cognizant of the magical underworkings of the database. For example:

- A book jacket image appears in the record because an ISBN was recorded in the metadata;
- A needed book can be borrowed from a library in Beijing because MARC records enable sharing of records in an international electronic union catalog;
- all the resources in a discipline or subject area can be perused because Library of Congress subject headings and classification have been added to the metadata; and
- all the works by a favorite author can be instantly called up because a standard authorized heading was used for the author’s name.

Metadata makes connectivity, interoperability, searchability, accessibility, and findability HAPPEN.

Types of Metadata

Whatever its purpose, a database runs on quality, standardized metadata that comes in a number of types. First, there is descriptive metadata that reports the intellectual content of a resource and aids in the discovery, identification, evaluation, collocation, and selection of resources. Then, there is technical metadata that describes information about creation and revision of digital objects, including resolution, compression, and pixel dimensions, that may be needed later for preservation purposes. Third, there is structural metadata that defines the relationships between multiple digital files. This is the “glue that binds a compound digital object together” (A Framework of Guidance for Building Good Digital Collections; Digital Library Forum) located at http://chnm.qmu.edu/digitalhistory/links/pdf/chapter3/3.18.pdf, and is used, for example, to synchronize audio with text, or facilitate navigation through an e-book. Finally, there is administrative metadata that facilitates management of information resources and records information about provenance, history, and ownership, and intellectual property rights.

Continued on page 2
Metadata
Continued from page 1

Components of Metadata
Wait, we’re not yet through our explanation! Besides the types just noted, metadata has three main components: syntax, semantics, and standards. As in language, metadata syntax, or encoding, defines the rules for construction of metadata “sentences.” Examples of syntax include Machine Readable Cataloging (MARC), a numerical encoding which enables one to go online to determine a library’s holdings, and Extensible Markup Language (XML), a “human readable” or language-based encoding which allows Web publishing, electronic data exchange, and portable, reusable metadata. XML is used in personal digital assistants, cellular phones, and automatic phone banking—and will be used in the library catalogs of the future.

In semantics, by contrast, we find the meaning of semiotic markers—in a metadata scheme as in language. But, because the word “chair” can refer to the piece of furniture or to the person presiding over a committee, a metadata system requires its third element, standards, which fix meanings that would otherwise—as in actual language—be unfixed, subjective, and contextual. Standards make possible the exchange of information by making metadata records compatible with each other and aiding interoperability between databases. There are standards for metadata element sets or schemes, element content, controlled vocabularies, and encoding.

Metadata Schemes
Because of the need for differences and levels of complexity in semantics for describing different types of resources, several different, but standardized metadata schemes have been developed; some are geared to specific disciplines and purposes. Among the most common are the:
- Visual Resources Association Core (VRA), used for describing cultural objects and works of art;
- Encoded Archival Description (EAD), for describing archived collections;
- Text Encoding Initiative (TEI), which facilitates the description and marking up of texts; and, most prominent, the
- Dublin Core, an all-purpose metadata scheme that, used in its simple or qualified forms, can integrate many different formats, including maps, images, and texts. In its simplest form, Dublin Core is a “lowest common denominator” scheme that facilitates system-to-system operability.

The original purpose of the Dublin Core was to organize the Web. Back in 1995, it was thought that the Web could be organized like a library if Web site creators would assign access points, descriptors, and subject headings to their content so that it could be located more easily. Web site creators did not have the motivation to catalog their Web sites, but museum curators, librarians, and visual arts librarians adopted the Dublin Core and were instrumental in its development and significance as a key component of the semantic Web.

WMU Metadata Projects
The Dublin Core metadata has already been used to describe the Caroline Bartlett Crane Collection and the Ward Morgan Photographic Collection (see Gatherings, No. 35, Fall 2004, p. 5) at WMU. Caroline Bartlett Crane was an early 20th century social activist. She designed an award-winning home, still standing in Kalamazoo, which was one of the first focusing on family spaces and convenience for women. The digitized collection includes blueprints of the house, photographs, and letters from Presidents Calvin Coolidge and Herbert Hoover congratulating her on winning the Better Homes in America 1924 Model Home contest. Dublin Core metadata is used to describe the objects, and, just as importantly, to provide a wide audience with Web access to a rare and important collection that is otherwise difficult to view.

Ward Morgan was a 20th century Kalamazoo commercial photographer. The digitized collection includes hundreds of photographs of businesses, homes, and schools in the Kalamazoo area. Using the metadata, the researcher can look for specific known businesses and buildings, neighborhoods, types of buildings, objects or people in the photographs, such as “women,” “soda fountain,” or “automobile dealers.” In addition, all photographs “about” certain topics may be gathered using browseable categories including education, advertising, and business products.

Two additional and exciting metadata projects in the works will be of great benefit to Western faculty and students. The first involves tens of thousands of slides featuring art objects that are in the process of being digitized and made available electronically for classroom use. VRA is the metadata scheme chosen for this project. VRA includes a set of elements, which may be encoded in XML, that is designed to describe digitized images of art objects, analog slides or photographs of the objects, and the physical objects themselves. Because of the specific art-related VRA elements used in the cataloging of the slides, the researcher will be able to bring together from this collection the works of art by a certain artist or architect, from a specific time period or genre, and even all works created using a particular technique or material.

As described in Sharon Carlson’s article, also in this issue of Gatherings, Western Michigan University recently received a grant of $95,000 to digitize and provide Web access to eight Civil War diaries currently housed in the Archives and Regional History Collections. Each diary is in the process of being digitized so that the actual pages may be viewed electronically from anywhere, and also transcribed into an easy-to-read format. The metadata scheme chosen to describe, provide access to, and preserve these valuable pieces of American history was TEI encoded in XML. Using TEI provides a way, not only to create an item-level record for each diary similar to a library catalog record, but also to “mark up” the transcribed text. Places in the text will be marked for retrieval where particular people, places, battles, military units, and topics are mentioned.

Thanks to metadata, information has indeed escaped its containers. Information resources, deteriorating, hidden, and remote, are rediscovered, shared, and preserved. Threatened cultures and histories, muted voices that have been long lost, are given a global stage. Metadata, the magic behind the scenes, makes it possible.

Books as Art

By Pam Rups, Coordinator, Instructional Technology Center, OIT

People feel so empowered when they make a book. I’ve seen that reaction when I teach and also felt it in my own experience. Because most of us tend to buy everything ready-made—our clothes, our food, our homes, our toys, and, especially our printed books—most of us have lost the ability to consider making something from scratch—unless the item is available in a kit or is some “thing” simple to put together. As a result, when someone makes a book for the first time, he or she feels the magic of creation.

On the other hand, the physical structure of books seems uninteresting to those who stop to think about them. I know very few people, especially in today’s electronic world, who have ever given any thought as to how a book is put together. Although most of us still use books almost daily: for reference, for study, for fun, and for training, we seem not to care about the package itself. If you ask people how this object, this book that