



2-2007

## Digitization Projects and Metadata

Sheila A. Bair

Western Michigan University, sheila.bair@wmich.edu

Pam Cowart

Western Michigan University, prebarcak@yahoo.com

Follow this and additional works at: [https://scholarworks.wmich.edu/library\\_pubs](https://scholarworks.wmich.edu/library_pubs)



Part of the Library and Information Science Commons

---

### WMU ScholarWorks Citation

Bair, Sheila A. and Cowart, Pam, "Digitization Projects and Metadata" (2007). *University Libraries Faculty & Staff Publications*. 15.

[https://scholarworks.wmich.edu/library\\_pubs/15](https://scholarworks.wmich.edu/library_pubs/15)

This Article is brought to you for free and open access by the University Libraries at ScholarWorks at WMU. It has been accepted for inclusion in University Libraries Faculty & Staff Publications by an authorized administrator of ScholarWorks at WMU. For more information, please contact [wmu-scholarworks@wmich.edu](mailto:wmu-scholarworks@wmich.edu).





# Against the Grain

*“Linking Publishers, Vendors and Librarians”*

ISSN: 1043-2094

## Digitization Projects and Metadata

by **Sheila Bair** (Metadata Cataloger, University Libraries, Western Michigan University, Kalamazoo, MI 49008-5353; Phone: 269-387-5160) <bair@wmich.edu>

and **Pam Cowart** (Head, Technical Services, Western Michigan University, Kalamazoo, MI 49009) <prebarcak@wmich.edu>

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 alphanumeric 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? One imagines more pages than can be bound, more books than can be shelved. More broadly, one imagines what librarians face as knowledge explodes and presents greater and greater challenges to those charged with storing or retrieving it. Now more than ever, librarians must think outside the old, linear, analog, bricks-and-mortar box. This is where we encounter metadata, which makes possible the finding, linking, and sharing of information on a granular level regardless of its format.

Notice that one says “metadata makes,” “metadata is.” The word “data,” like the word “agenda,” may have begun life as a plural, but already it sounds strange — indeed, pedantic — to say, “the data are.” So: what is metadata? The Greek prefix tells the story: it means beyond, about, sometimes behind — as in metaphysics, metabolism, metempsychosis, and so on.

Beyond data, then. Data about data. Indeed:

### Data Behind Data

Metadata is the data that, working quietly behind information, makes it accessible and coherent, thereby enabling people and systems

to do smarter things. We are not, then, talking about randomly accumulated data, nor do we mean structured information in any ordinary sense. Ordered according to one of several standard schemes, metadata is, as **Priscilla Caplan** explains, information that describes or identifies other information — or another information source.<sup>1</sup> **The National Science Digital Library** further defines metadata as “structured, standardized descriptions of resources, whether digital or physical, that aid in the discovery, retrieval and use of those resources.”<sup>2</sup> Such information about information becomes increasingly valuable in a knowledge economy: 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.



Exploding by the nanosecond, information threatens to overwhelm the frail bark of our capacity to order it. But information seekers are seldom cognizant of the metadata behind 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 effectuates connectivity, interoperability, searchability, accessibility, and findability.

### Types of Metadata

Whatever its purpose, a database runs on quality, standardized metadata, which comes in a number of types. *Descriptive* metadata aids in the discovery, identification, evaluation, collocation, and selection of resources. *Technical* metadata describes information about creation and revision of digital objects, including resolution, compression, and pixel dimensions — information that may be needed later for migration. *Structural* metadata defines the relationships between multiple digital files. As it “relates the pieces of a compound object together,” it can synchronize audio with text or facilitate navigation through an eBook.<sup>3</sup> *Administrative* metadata, finally, facilitates management of information resources and records information about provenance, history, ownership, and intellectual property rights.

### Components of Metadata

We further characterize metadata in terms of 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)**, an alphanumeric encoding that enables one to go online to determine a library’s holdings, and **Extensible Markup Language (XML)**, a “human readable” or language-based encoding that allows Web publishing, electronic data exchange, and portable, reusable metadata. A feature of personal digital assistants, cellular phones, and automatic phone banking, **XML** will figure importantly 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 where the word “chair” can refer to the piece of furniture or to the person presiding over a committee. Thus a metadata system requires a third and final component, standards, which fix meanings that would otherwise — as in actual language — be unfixed, subjective,

*continued on page 18*

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. The most common of those geared to specific disciplines and purposes include:

- **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,
- **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 Website creators would assign access points, descriptors, and subject headings to their content so that it could be located more easily. Website creators did not have the motivation to catalog their Websites, 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.

### The Articles in This Issue

Collection development increasingly features digitization of hidden resources, unique collections, and rare materials. But digitization involves more than just scanning items in some Web-friendly format. It involves metadata, the key to making a digital collection easily searchable, compatible with local, consortial, and even global systems — and accessible into the future.

Contributors to this issue of *Against the Grain* emphasize the importance of coordinating with catalogers from the beginning of any digitization initiative. Doing so will save much backtracking and associated expense later. Thus collection decision makers and metadata catalogers/specialists should continue to forge strong relationships to bring the best product to the user.

Traditionally, collections librarians have chosen materials represented in the catalog by a **MARC** record. Raised on the ISBDs, firmly married to the content standard **AACR2**, and happily housed in your local ILS, **MARC** is

## against the grain people profile

Cataloger/Metadata Cataloger, University Libraries  
Western Michigan University, Kalamazoo, MI 49008-5353  
Phone: (269) 387-5160 <bair@wmich.edu>

### Sheila Bair, MLIS

**BORN & LIVED:** Born in Kalamazoo, Michigan, for past 17 years I have lived in the country just west of Kalamazoo in Van Buren County, Michigan.

**EARLY LIFE:** Raised by two professional ballroom dance instructors (one an amateur actress) — never a dull moment.

**FAMILY:** Married with three children and two very sweet grandchildren.

**EDUCATION:** B.A. in Communications from **Western Michigan University**. MLIS from **University of Wisconsin-Milwaukee**.

**FIRST JOB:** Worked in the university library. I did everything, circulation, shelf-reading, sent out overdue notices, typed up and filed catalog cards.

**PROFESSIONAL CAREER AND ACTIVITIES:** Cataloging/Metadata Librarian at **Western Michigan University**.

**IN MY SPARE TIME I LIKE TO:** Ride my bike through the vineyards in Michigan's beautiful countryside, read a good detective mystery.

**FAVORITE BOOKS:** *Detective Rebus* novels by **Ian Rankin**, **Charles Dickens**, **Jane Austen**.

**PET PEEVES/WHAT MAKES ME MAD:** Negativity, mediocrity.

**PHILOSOPHY:** You can learn something from every person and every situation.

**HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS:**

From the cataloging/metadata perspective, a dramatic and revolutionary change is taking place. Cataloging rules, theory and practice, and catalogers themselves will continue to be fully integrated into the electronic environment. The job title “cataloger” may not exist in five years, with the title “metadata creator” or some other term, as yet unimagined, taking its place. 🐾



a well established schema. Those collecting standard resources rarely had to wonder, “How will we provide access?” When selecting resources for digitization, however, collection development principles must be augmented by answers to a host of questions. How will digital assets be preserved? What schema will be used to describe them? What system will house them?

In this issue, we hope to answer these questions and others. First of all, **Jody Perkins** will give a conspectus of the essential matters that planners of a digital project need to take into consideration. Her excellent checklist includes sixteen vital points to consider when evaluating a collection. She discusses metadata design, choosing schemas and standards, and documenting decisions through the use of a data dictionary.

Reflecting further on schema selection, **Jeffrey Beall** enumerates twelve points of comparison to help one decide which of the many schemas available best suits one's digital project. He addresses such concerns as interoperability, granularity, proven success, and level of community or domain specificity.

Next, a pair of case studies: **James Bradley** discusses the efficacy, for a digital image col-

lection, of **CONTENTdm** and **Dublin Core**; and **Jen Wolfe** and **Mark F. Anderson** review the difficulties and decision-making involved in opting for **DigiTools** and **METS** to provide access to a collection of science fiction fanzines. These case studies cover crosswalking, the viability of existing schemes, copyright issues, and decisions about the depth and extent of metadata needed.

Finally, **Arwen Hutt**, **Trish Rose-Sandler**, and **Bradley D. Westbrook** share one library community's successful approach to metadata preservation, a hot topic that the digital library community must concern itself with, especially complex problems of long-term usability. In their article, they describe creation of a digital asset management system that, ingeniously wrapping **MODS** in **METS**, converts different types of metadata from many diverse projects into one interoperable and manageable schema.

These essays offer a wealth of insight into some of the most important electronic resources issues currently facing collection development. As we digitize our unique holdings, preserve items in jeopardy, or offer our most popular collections to the broadest user

*continued on page 20*

base, we would do well to keep in mind that the important decisions are made at the beginning of the collection digitization project and are mission critical to current and future plans for interoperability

Thanks to metadata, information has indeed escaped its containers. Deteriorating, hidden, and remote information resources are rediscovered, shared, and preserved. Muted voices, threatened cultures, whole histories that have long been buried find themselves at a global stage-center. Metadata makes it possible. 🌱

#### Endnotes

1. **Priscilla Caplan**, *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003).
2. The **National Science Digital Library**, *NSDL Metadata Primer*, <http://metamanagement.comm.nsdlib.org/overview2.html#what>.
3. **NISO Framework Advisory Group**, *A Framework of Guidance for Building Good Digital Collections*, 2nd ed. (2004): 27, <http://www.niso.org/framework/Framework2.pdf>.

## against the grain people profile

Head, Technical Services Librarian  
Western Michigan University, Kalamazoo, MI 49009  
<prebarcak@wmich.edu>

### Pam Cowart

**BORN & LIVED:** Born in Texas, I have lived in Louisiana, South Carolina, Iowa, Michigan, and Italy.

**EARLY LIFE:** Only child.

**FAMILY:** Recently married (formerly Pam Rebarcak), two stepchildren.

**EDUCATION:** BA **Louisiana Tech University**; MLS **Louisiana State University**; MPA **Lamar University**.

**FIRST JOB:** Lifeguard at a summer camp.

**FIRST PROFESSIONAL JOB:** Serials Cataloger at **University of South Carolina**.

**PROFESSIONAL CAREER AND ACTIVITIES:** Cataloging, Serials, Acquisitions, Technical Services, Social Sciences Bibliographer, and, as an Account Executive with LexisNexis, marketing, sales, and training. Served on numerous ALA and ALCTS Committees.

**IN MY SPARE TIME I LIKE TO:** Wear out treadmills, go to movies and museums, travel (most recently to India).

**FAVORITE BOOKS:** *Lucky Jim*, *Motoring with Mohammed*, *A Severed Head*, *The Moviegoer*, *Kate Vaiden*.

**PET PEEVES/WHAT MAKES ME MAD:** Slow drivers in the passing lane.

**PHILOSOPHY:** Life is short – live it.

**HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS:** Librarians will do what they've always done – rise above the confusion and make sense of it all. 🌱

