


Fall 2006

## Metadata Basics: A Literature Survey and Subject Analysis

Nicole Mitchell  
nmitchell79@gmail.com

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/seln>

 Part of the [Databases and Information Systems Commons](#), and the [Library and Information Science Commons](#)

---

### Recommended Citation

Mitchell, Nicole (2006) "Metadata Basics: A Literature Survey and Subject Analysis," *The Southeastern Librarian*: Vol. 54 : Iss. 3 , Article 6.

Available at: <https://digitalcommons.kennesaw.edu/seln/vol54/iss3/6>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in The Southeastern Librarian by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact [digitalcommons@kennesaw.edu](mailto:digitalcommons@kennesaw.edu).

## Metadata Basics: A Literature Survey and Subject Analysis

Nicole Mitchell

Nicole Mitchell is a Graduate Student and IMLS Fellow at the School of Library and Information Studies at the University of Alabama. She can be reached at: [nmitchell79@gmail.com](mailto:nmitchell79@gmail.com)

### Introduction

Librarians today are wrestling with an ever-changing digital environment. In some way or another, we must all adapt to new technologies, skills, and ways of thinking. What comes to mind when you hear the word “metadata?” Is it intimidating? Do metadatists and catalogers explain the term adequately? While this article by no means captures all there is about metadata, it is intended to provide librarians with a basic understanding of what is involved in metadata work.

While a number of articles about metadata have been written for specialized journals, few have been written for the library community at large. Many librarians who are non-catalogers have only a vague notion of metadata. Unfortunately, there is not just one definition for the term metadata but rather a variety of definitions and explanations. While the term *metadata* first emerged in the computer world in the 1960s, it did not appear in the library community until the mid 1990s. For librarians, metadata is very simply data about data. Some, including former American Library Association President Michael Gorman and Tom Delsey of the National Library of Canada, consider metadata to be “cataloging done by men.”<sup>1</sup> Others think of it as ‘dumbed-down cataloging,’ since authors often create the metadata for a work in addition to the work itself. In his “Digital Libraries” column for *Library Journal*, Roy Tennant even referred to it as “cataloging by those paid better than librarians.”<sup>2</sup>

---

<sup>1</sup>Karen Coyle, “Understanding Metadata and Its Purpose,” *The Journal of Academic Librarianship* 31, no.2 (March 2005), 1.

<sup>2</sup>Roy Tennant, “Metadata As If Libraries Depended on It,” *Library Journal* 127, no.7 (April 15, 2002), 32.

### What is Metadata?

In a recent volume of *Cataloging and Classification Quarterly*, editor Richard P. Smiraglia states that the main purpose of the articles is to “search for a definition.”<sup>3</sup> While traditional library cataloging is also a form of metadata, many feel that the term is applicable when describing only electronic or digital resources.<sup>4</sup> In this vein, the International Federation of Library Associations defines metadata as “any data used to aid the identification, description and location of networked electronic resources.” The American Library Association’s Committee on Cataloging: Description and Access developed even another definition: “structured, encoded data that describe characteristics of information-bearing entities to aid in their identification, discovery, assessment, and management of the described entities.”<sup>5</sup> The National Information Standards Organization (NISO) provides us with a much more precise definition of metadata: “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.”<sup>6</sup> None of these definitions, however, really explain what metadata is. Metadata is, quite simply, any type of formal description of a resource, regardless of format.

After looking at the various definitions of metadata, one may ask why it is even necessary to distinguish metadata work from traditional

---

<sup>3</sup>Richard P. Smiraglia, “Introducing Metadata,” *Cataloging and Classification Quarterly* 40, no.3/4 (2005), 2.

<sup>4</sup>NISO, *Understanding Metadata*. Bethesda, MD: NISO Press. Available at <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

<sup>5</sup>Library of Congress, *Metadata Basics*, Available at <http://www.loc.gov/catworkshop/readings/metadatabasics/>

<sup>6</sup>NISO, “Understanding Metadata.”

library cataloging. Just as there are a number of definitions for the term metadata, there are also several types of metadata, including administrative and technical, structural, preservation, and descriptive. Administrative and technical metadata consists of information about how the resource was created, format and file type, and access restrictions. Structural metadata is that which indicates how a resource is arranged, i.e., number of pages or chapters. Preservation metadata includes any information necessary to archive and preserve the resource. For the purposes of this paper, however, we will only be concerned with descriptive metadata. What, then, are the functions of descriptive metadata? Creation of descriptive metadata is essentially what catalogers do everyday. The primary function, as with any formal resource description scheme, is to describe resources so that users can search for and locate them. Descriptive metadata is also useful in organizing and linking resources as well as sharing data across repositories.

### Metadata Schemas

Examples of frequently used metadata schemas include Dublin Core, Metadata Encoding and Transmission Standard (METS), the Text Encoding Initiative (TEI), Metadata Object Description Schema (MODS), and Encoded Archival Description (EAD). To achieve the functions of descriptive metadata, a metadata schema (simply a list of elements) must be interoperable; in other words, it must be able to exchange information with other systems. Just as online public access catalogs share information through machine-readable cataloging (MARC) records, metadata schemas must also be able to talk to each other across formats. The interoperability of metadata schemas allows users to search across different systems more efficiently and allows for a more seamless transition between these different systems. With the Z39.50 protocol and others, users are able to search among a variety of resources no matter how the resources were organized or described. Interoperability also aids in metadata harvesting—the retrieval of metadata records from multiple repositories. Harvesting records

can be difficult when metadata schemas are not compatible and cannot be translated into a universal record format. Good content is often lost when the information in one metadata schema cannot be converted into another. For instance, one may want to convert a MARC record into a simple Dublin Core record (more on this later). While a MARC record has numerous fields, a simple Dublin Core record has only 15. There is nowhere in the Dublin Core schema to place the content from the extra MARC fields. That extra content is misplaced if it is put into a single Dublin Core element. In turn, this affects the indexing of Dublin Core metadata records.

In addition to being interoperable, a good metadata schema must also be flexible enough to be used in a number of different information organizations and communities. Drawing on information from Lois Mai Chan, author Rosemary Aud Franklin concurs that the flexibility of differing metadata schemas is essential to “accommodate the need for different degrees of depth and different subject domains” that may be used in different libraries, museums, or archives.<sup>7</sup> To ensure both flexibility and interoperability of metadata schemas, developers suggest that a controlled vocabulary, rather than natural language, be used to describe resources.

### Dublin Core

Perhaps the most common metadata schema used throughout the library community today is Dublin Core. Developed in 1995, the Dublin Core Metadata Initiative is the result of a workshop held by OCLC and the National Center for Supercomputing Applications (NCSA). From the start, Dublin Core was aimed at non-professionals. It was intended to be simple enough that individual authors or creators could describe their own material and web pages. Because of this, Dublin Core is very broad in scope, attempting to meet the metadata needs of several communities. Dublin Core can be used for a variety of purposes such as adding metadata to web resources. It can also be used as a common meeting ground for more complex metadata schemas.

---

<sup>7</sup>Franklin, Rosemary Aud. “Re-Inventing Subject Access for the Semantic Web,” *Online Information Review* 27, no.2 (2003), 96.

A basic Dublin Core record has just fifteen elements:

- a. Title
- b. Creator
- c. Subject
- d. Description
- e. Publisher
- f. Contributor
- g. Date
- h. Type
- i. Format
- j. Identifier
- k. Source
- l. Language
- m. Relation
- n. Coverage
- o. Rights

Though Dublin Core only utilizes fifteen elements, some, like date, can be “qualified” to make them more specific. Records with qualified elements are known as Qualified Dublin Core, while records without such qualified elements are known as Unqualified or Simple Dublin Core. In addition to being able to qualify elements, all Dublin Core elements are optional, can be repeated, and can be arranged in any order. There are virtually no rules for creating Dublin Core records. Because of this, interoperability between metadata schemas is often a problem. For instance, an author’s name may be entered “first name last name” or “last name, first name.” While Dublin Core does recommend using a controlled vocabulary for certain elements, it is not required. This is only a recommended best practice and is optional. As Priscilla Caplan illustrates, Dublin Core “allows the advantage of some standardization while giving project designers the leeway to identify data elements and guidelines that are meaningful to them.”<sup>8</sup>

### LCSH and Subject Analysis

When the notion of metadata first emerged in the library world in the early 1990s, it was obvious that provisions for subject access would be necessary. With so many communities using

---

<sup>8</sup>Caplan, Priscilla. *Metadata Fundamentals for All Librarians*. (Chicago: ALA, 2003), 85.

different vocabularies and terminologies, metadata schemas cannot prescribe the use of one specific vocabulary. While the majority of academic and research libraries have long since used Library of Congress Subject Headings (LCSH), museums, archives, and even other libraries have their own specialized thesauri and classification schemes such as the Art and Architecture Thesaurus and Medical Subject Headings. With such varying vocabularies, a consensus was needed to determine how metadatists should approach subject access. Individual institutions looked to professional library organizations for support.

In 1997, the American Library Association’s Association for Library Collections and Technical Services (ALCTS) division formed the Subcommittee on Metadata and Subject Analysis in order to “identify and study the major issues surrounding the use of metadata in the subject analysis and classification section of digital resources.”<sup>9</sup> The group focused particularly on the Dublin Core metadata schema. In July 1999, the group issued its final report entitled “Subject Data in the Metadata Record: Recommendations and Rationale,” which addressed many of the key issues of subject access. The group also stated that the use of multiple vocabularies should be accommodated. Utilizing an existing vocabulary would be helpful in achieving semantic interoperability. The subcommittee recommended using LCSH or the Sears List of Subject Headings for a general vocabulary that would cover all subjects. While the “level of specificity of LCSH would be a good basis,”<sup>10</sup> the subcommittee recommends combining a controlled vocabulary with subject-related keywords to enhance search retrieval. In addition to aiding the search process, using an existing vocabulary would ensure compatibility with “the enormous store of MARC records in OPACS.”<sup>11</sup>

If we are to employ controlled vocabularies like LCSH in metadata records, it will be helpful to

---

<sup>9</sup>Association for Library Collections and Technical Services (ALCTS), “Subject Data in the Metadata Record: Recommendations and Rationale,” A Report from the Subcommittee on Metadata and Subject Analysis.

<sup>10</sup>Ibid.

<sup>11</sup>Ibid.

look at them more closely. Widely used in the library community, LCSH consists of a large and non-specialized vocabulary. LCSH are especially rich because they cover a variety of subjects. They also offer pre-coordinated (meaning the terms were combined when they were established) terms, show the semantic relationships between terms, and have dependable authority control. LCSH are also extremely versatile, expanding and incorporating new terms as needed. Despite such a rich vocabulary, the LCSH are not especially intuitive or easy to use. While catalogers tend to assign very specific subject headings, typical users search only for general terms or keywords. Further, LCSH are not suited for online search engines due simply to their complexity. How then will LCSH fit into emerging metadata schemas?

In their 2000 article, Lois Mai Chan and Theodora Hodges tackle what the millennium holds for LCSH and how LCSH would “adapt to the multifarious environment.”<sup>12</sup> After a discussion of the development of LCSH, the authors conclude that the system must change if it “is to play an important role in subject access to information.”<sup>13</sup> Chan and Hodges determined that a possible solution is to simplify the LCSH string by developing a system that is post-coordinate (where terms are put together during the search process rather than in the index) and separated into more manageable parts. Franklin agrees that a faceted approach would “support the need to describe content that is not easily collapsed into rigid hierarchies.”<sup>14</sup> Chan and Hodges go on to outline four advantages for transitioning to a faceted, post-coordinate approach to subject data in metadata records:

1. “A postcoordinate approach is more adaptable and amenable to changes in the web environment. It will not require the extensive training necessary to apply LCSH according to current policies and procedures;

2. An online thesaurus based on faceted principles is easier to display and for non-catalogers or non-librarians to use;
3. A postcoordinate subject vocabulary is compatible in structure and syntax [how words are arranged] with most other controlled vocabularies; and,
4. A postcoordinate single-term approach is more amenable than full subject strings for mapping to other controlled vocabularies, to other languages, or to classification schemes such as DDC.”<sup>15</sup>

## FAST

Along these lines, OCLC has developed FAST—Faceted Application of Subject Terminology—for use in Dublin Core metadata records. A new approach to subject vocabulary, FAST was based on LCSH and was designed specifically for an online environment, consisting of post-coordinated and faceted terms. Essentially, FAST breaks up the LCSH strings into four more manageable parts or facets. These facets consist of topic, geographic, form, and period. A typical LCSH would look like this: “Georgia—History--Civil War, 1861-1865—Battlefields--Guidebooks.” FAST would break down the subject string into the following facets:

Georgia (Geographic)  
 History—Civil War, 1861-1865 (Topic)  
 1861-1865 (Period)  
 Battlefields (Topic)  
 Guidebooks (Form)

By simply altering an existing vocabulary rather than creating a new one, FAST will be compatible with LCSH and the subject data in the majority of MARC records. Since FAST is based on LCSH, the “automated conversion of LCSH to the new [system] is possible.”<sup>16</sup> Authors Edward T. O’Neill and Lois Mai Chan also suggest that maintaining the FAST system will be less costly since any changes to LCSH can be

<sup>12</sup>Chan, Lois Mai and Theodora Hodges. “Entering the Millennium: A New Century for LCSH,” *Cataloging & Classification Quarterly* 29, no.1/2 (2000). 225

<sup>13</sup>Chan and Hodges, 229.

<sup>14</sup>Franklin, 100.

<sup>15</sup>Chan and Hodges, 232.

<sup>16</sup>O’Neill, Edward T. and Lois Mai Chan. “FAST (Faceted Application of Subject Terminology): A Simplified Vocabulary Based on the Library of Congress Subject Headings,” *IFLA Journal* 29 (2003), 337.

automatically included in FAST. They go on to say that

“by separating syntax from semantics, the application process can be simplified while retaining the richness of vocabulary in LCSH thus making the schema easier to use and maintain. Furthermore, with the simplified syntax and application rules, computer technology can be used to greater advantage in both the assignment and the maintenance of subject data as well as in subject authority control.”<sup>17</sup>

As a post-coordinated vocabulary, FAST will be more usable for people with minimal training and experience since Dublin Core was intended to be used by non-catalogers. Among the many benefits of FAST, O’Neill and Chan also believe that the new vocabulary will “be able to accommodate different retrieval models [and] facilitate mapping of subject data and cross-domain searching.”<sup>18</sup>

### Who Creates Metadata?

After looking at metadata and subject access approaches, one may wonder who is actually doing metadata work. Is it professional librarians, untrained non-catalogers, archivists, students, subject specialists? Because Dublin Core was designed for non-professionals, “the assignment of metadata in Dublin Core requires less training, time, and reliance on supplemental resources than traditional cataloging practices, making it more feasible to use non-professional labor in metadata construction.”<sup>19</sup> At the *Civil Rights in Mississippi Digital Archive*, metadata is largely created by students in the School of Library and Information Science at the University of Southern Mississippi. While students choose initial subject headings from a controlled vocabulary, a metadata librarian “proofreads the records to correct errors and to ensure appropriate subject heading assignment.”<sup>20</sup>

While a number of libraries are engaged in metadata activities, it appears, in some cases, that little emphasis is placed on subject access. In “Lessons Learned from the Illinois OAI Metadata Harvesting Project,” Timothy W. Cole and Sarah L. Shreeves found that “academic libraries—institutions with strong traditions of descriptive cataloging—made surprisingly infrequent use of the ‘subject’ element” and “few used standard controlled vocabularies.”<sup>21</sup> When looking at metadata records from seven academic libraries across the United States, Cole and Shreeves were dismayed to discover that only about fifteen percent of the metadata records contained the subject element. They concluded that metadata creators in academic libraries clearly “are not following the same content creation rules they use for creating catalog records describing their print collections.”<sup>22</sup> The question to be asked now is “why?”

### Training and Education

Following the discussion of both metadata and subject access approaches in metadata records, it is interesting to examine how emerging metadata schemas impact education and training programs. In November 2000, the Library of Congress held the Bicentennial Conference on Bibliographic Control for the New Millennium since “competencies in cataloging and metadata have become critical for library information professionals to be effective and competitive.”<sup>23</sup> As a result of this conference, five action plans were proposed to maintain bibliographic control of web resources. Action 5 is concerned with training and education programs needed to provide “students with core competencies in technical services and management skills [and] produc[e] creative and resourceful catalogers.”<sup>24</sup> The goal was to garner more participation from new library and information science professionals in developing metadata standards.

---

<sup>17</sup>Ibid., 337.

<sup>18</sup>Ibid., 342.

<sup>19</sup>Graham, Suzanne R. and Diane DeCesare Ross. “Metadata and Authority Control in the Civil Rights in Mississippi Digital Archive,” *Journal of Internet Cataloging* 6, no.1 (2003), 35.

<sup>20</sup>Ibid., 36.

---

<sup>21</sup>Cole, Timothy W. and Sarah L. Shreeves. “Lessons Learned from the Illinois OAI Metadata Harvesting Project,” in *Metadata in Practice*, (Chicago: ALA, 2004), 175.

<sup>22</sup>Ibid., 184.

<sup>23</sup>Hsieh-Yee, Ingrid. “Cataloging and Metadata Education in North American LIS Programs,” *Library Resources & Technical Services* 48, no.1 (Jan., 2004), 59.

<sup>24</sup>Ibid., 60.

As part of the group working to “improve and enhance curricula in library and information science schools,”<sup>25</sup> lead researcher Ingrid Hsieh-Yee conducted a survey of all ALA-accredited schools. She found that fewer programs required cataloging courses, relying instead on introductory courses to broach the topics of cataloging and metadata. Only three of the respondents said that they even discuss the relationship between cataloging and metadata. As a result, students are graduating with insufficient knowledge and preparation for either cataloging or metadata positions. Hsieh-Yee points out, however, that, “for aspiring catalogers and metadata specialists alike, competencies in cataloging and metadata are essential.”<sup>26</sup>

Since library schools are not sufficiently preparing students to be catalogers or metadatists, one might expect the American Library Association, for example, to take the lead. In 2006, the Committee on Education, Training, and Recruitment for Cataloging of the Cataloging and Classification Section of ALCTS prepared a document entitled “Training Catalogers in the Electronic Era: Essential Elements of a Training Program for Entry-Level Professional Catalogers.” Though this document was “intended to assist in the training of beginning professional catalogers,” it is particularly telling that there is absolutely no mention of metadata anywhere in the text. Rather than a training manual, “Training Catalogers in the Electronic Era” serves as more of an orientation manual.

In Hsieh-Yee’s “Cataloging and Metadata Education: A Proposal for Preparing Cataloging Professionals of the 21st Century” which she submitted to the ALCTS/ALISE Task Force, she outlines five critical steps that will help educators better prepare their students. The proposed actions include:

1. Publicizing the levels of expertise and competencies of LIS educators and practitioners

2. Create “Metadata Basics” package of resources and tools
3. Create a listserv to discuss cataloging and metadata education
4. Create web clearinghouse on cataloging and metadata education
5. Hold conference on effective teaching strategies

Begun in 2003, work is still progressing. Hsieh-Yee’s proposal called for an update on the progress and future plans at Mid-Winter ALA in January 2006.

So many definitions of metadata often find non-cataloging librarians grappling with what this all means. New library professionals are especially at a disadvantage. Though they may have the technical and computer skills to do metadata work, they often lack the knowledge of and theory behind traditional library cataloging. So, for new and aspiring librarians, the world of metadata can seem confusing and chaotic. We no longer have traditional standards, such as Anglo American Cataloging Rules (AACR), to employ with all metadata schemas. Instead, there are “best practices guidelines.” Rather than providing explicit instructions, “best practices guidelines” simply make recommendations. This can leave newer librarians wondering which “best practice” they should employ. Determining which thesauri or schema to use can be a daunting task as well. How do you decide when to use FAST instead of LCSH? How do you determine which metadata schema to use?

With so many job advertisements requiring metadata skills, it is apparent that this trend will continue as more and more resources become digitized. A number of questions come to mind after perusing these ads. For one, how do we prepare applicants for these positions—through formal programs or on-the-job training? Should students be getting these technical skills in library school? Should employers expect their future applicants to have these skills already? What about the continuing education needs of current catalogers? They also need training. The

---

<sup>25</sup>Library of Congress, Bibliographic Control of Web Resources: A Library of Congress Action Plan, Available at [www.loc.gov/catdir/bibcontrol/actionplan.pdf](http://www.loc.gov/catdir/bibcontrol/actionplan.pdf)

<sup>26</sup>Hsieh-Yee, 65.

skills needed for metadata work are often more technical than those needed for traditional cataloging. Where do current catalogers and metadatists turn for resources?

### **What Does the Future Hold?**

In a recent article in *Educause Review* entitled “Changing a Cultural Icon: The Academic Library as a Virtual Destination,” Jerry D. Campbell asserts that, “if librarians are involved at all [in the new library], it is already clear that their role with respect to metadata will be vastly different from their old cataloging role.”<sup>27</sup> Campbell claims that scholars were forced “to get into the digital library business in order to save, use, and manage their own data” because librarians lacked both interest and technical skills.<sup>28</sup> Is the situation Campbell presents the whole truth? Are librarians just not interested? Or do they just have nowhere to get the required technical skills necessary for metadata and digital library work?

---

<sup>27</sup>Campbell, Jerry D. “Changing a Cultural Icon: The Academic Library as a Virtual Destination,” *Educause Review* (Jan/Feb., 2006), 22.

<sup>28</sup>*Ibid.*, 22.

While a number of questions about metadata remain unanswered, there are places to turn for guidance. The following list of resources should prove helpful in bettering your understanding of metadata.

### **Useful Websites:**

*Metadata Basics:*

<http://www.loc.gov/catworkshop/readings/metadata/basics/>

*Understanding Metadata:*

<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

*Dublin Core Metadata Initiative:*

<http://www.dublincore.org/>

*Library of Congress Standards:*

<http://www.loc.gov/standards/>

*FAST:*

<http://www.oclc.org/research/projects/fast/>