Stephanie Girard
Nick Westbrook
Lydia Holland
Annie Horsman
ILS 501.01
12/1/12

Metadata: Present and Future Applications

Metadata, which is often referred to as data about data, is used to help describe various types of information, making it easier to find or use. In this paper, we will examine the different ways metadata is used for both digital and non-digital data, as well as how metadata is structured and used regarding syntax and semantics. Literature regarding the various schemes for recording metadata, such as Dublin Core and MODS, and TEI will be reviewed. In addition, this paper will describe the uses of metadata in different settings, including digital libraries, non digital libraries, and museums. The purpose of this paper is to provide an overview metadata in an attempt to provide context for the different schemas and format of metadata and why it is important to use in order to organize and locate information.

What is metadata?

NISO describes metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource." In the library environment, metadata is the term used for a formal scheme of resource description. There are three main types of metadata recognized by libraries and other institutions that have data management needs. These are descriptive metadata, structural metadata, and administrative metadata. Descriptive metadata describes the resource or information for purposes of identification. This can include element sets such as author or keywords. Structural metadata is meant to identify how objects or bodies of work are put together, and can include the table of contents, or page numbering. The last type of metadata, administrative, is meant to provide

information that would help manage a resource, such as creation dates, rights management, and archive management information (NISO, 2004). In a world where there is new technology and information created every day, it becomes increasingly important to use metadata as a means of controlling, managing, and organizing both physical and digital information.

What does metadata do?

One purpose, and probably the most important, is resource discovery. The way the metadata is presented (especially in the digital format) allows for cross-referencing. This means that a person looking at Renaissance Art can find images, books and articles about the same artist or person by following certain elements of metadata. Resource discovery is an important part of creating metadata because it makes researching a subject easier. The descriptive metadata is used specifically for this purpose because it provides a set of keywords and other elements to search for that make it easier to find the resources.

Another major purpose of metadata is interoperability. This, according to NISO Press's book *Understanding Metadata*, is described as "the ability of multiple systems, using different hardware and software platforms, data structures, and interfaces, to exchange and share data" (NISO, p.15, 2004). This means, that every metadata scheme is structured similarly enough, that the same piece of information can be represented in different metadata schemes. The title, for example can be expressed in both the Dublin Core (which is simply expressed as "title") and MARC 21 (expressed as "245 00$a"), where the title is the same for both. Searching for the title in both metadata systems yields the same result and should take you to similar documents.

The other main purpose behind metadata is digital preservation. Most current efforts to use metadata are focused on this purpose. When a document is digitized, it's important to keep track of it. Digital metadata is used to find the document after it's been placed into a digital

collection. This makes digitization of documents much easier and allows for the preservation of materials much later in the document's timeline.

Types of Metadata

There are three types of metadata: descriptive, structural and administrative.

Descriptive metadata includes elements such as author, title, keywords and other relevant data when searching for a document. This type of metadata is useful for resource discovery and identification of information. This describes a resource in terms of the content and type of the document.

Structural metadata describes the physical structure of the document itself, such as how a book is order (pages broken down into chapters, etc). Usually, structural metadata describes the intellectual or physical elements of a digital object (such as how many pages are in the document).

Administrative metadata includes the copyright information, as well as information relating to the digital preservation of the document. This information is different because it is designed with management in mind. This describes what type of file the data is, as well as how and when it was created. Within administrative metadata are two subsets of metadata: rights management (which protects the intellectual property writes of the document) and preservation (which contains the information needed to archive and preserve the resource).

Different Applications of Metadata

When considering metadata, there are many different systems (both physical and digital) that require the various systems and types. Digital libraries and different kinds of physical catalogs use metadata in different ways to achieve the same goal. In the case of catalogs, the main goal is to create a set of precise, directly related data points for an artifact or document.

For digital libraries, the goal is more to create a searchable database that can cross reference different documents, so research becomes more efficient and the resources are permanently archived for use of future researchers.

Catalogs are use in both libraries and various types of museums. In the case of libraries, catalogs can be used to organize information into an easily searchable system for easy access for patrons. A patron can search either a digital catalog (or an older, card catalog) and find resources relating to certain topics, authors or types of data without the libraries systems.

Museum catalogs keep track of artifacts in a specific museum and provide information about the specific pieces. The metadata needed for artifacts in museums help the museum staff organize collections, by grouping together time periods or specific artists within the museums. These catalogs are designed more for the museum staff than the patrons, but the information could be provided for displays (should the need arise).

The final type of metadata application is the use of metadata in digital libraries. Short of the digital documents, the metadata of these documents is probably the most important part of a digital library. This allows a user to search the library for specific subjects and find them without searching through the entire library. This makes researching a subject much easer.

Use in Libraries

When considering metadata in libraries, the purpose is focused primarily on resource discovery and data preservation. In academic world, metadata schemes assist in organizing and researching the resources at the disposal of students and staff. The metadata is used to create a searchable database to increase productivity and offer additional resources to the researcher. In the world of public libraries, its patrons can search similar databases and be led to similar articles as well as other sources of information (such as websites or databases) that provide additional

leads to new documents that might otherwise have been unnoticed (digital or otherwise). Considering the biography of an artist, actual documentation of the artist's work might not be something a person would think of (such as examining their paintings or sculptures.

Metadata can also be used for cataloging items into a structure that makes it easy to find. For example, all of the books containing a certain set of keywords are grouped together in a library, so if a person needs a certain subject, all of the books are together. Similarly, documents can be grouped by type, allowing for researchers to look for specific types of resources in one place (a patron of a library or museum could search the archives for magazines or newspapers pertaining to a certain date range and find them all together, connecting their research in that way).

Metadata Structure

Metadata can be created according to different schemes. Schemes consist of varying syntax, semantic, and content rules in order to describe certain types of information resources. The elements used to describe information are referred to as the semantics of the metadata scheme, while the values given to the elements are known as the content. NISO, a non-profit organization that maintains and publishes technical standards to manage information and metadata, describes the relationship between metadata schemes and semantics and content:

> Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title), representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary (NISO, p.2, 2004).

An example of the difference in semantics and content of metadata schemes can easily be found by looking at the user-friendly Dublin Core scheme and the numeral based tags used by the complex MARC 21 scheme. While Dublin Core is easily used by creating elements and tags that are easy to read, the MARC 21 scheme uses numerical tags and codes to describe information (Dutta, 2003).

Syntax is the final way metadata is structured. Syntax rules refer to how content and information should be encoded. Current schemes often include SGML (Standard Generalized Mark-up Language) or XML, an extended form of HTML that "allows for locally defined tag sets and the easy exchange of information," (NISO, p.2-3,2004). XML creation tools are becoming increasingly more widespread online, and therefore XML is used increasingly as the choice syntax for many different metadata schemes.

Metadata Schemes

Metadata schemes are generally understood to mean a classification or terminological system, which historically are used in library catalogues and databases. Most recently, metadata schemes have become increasingly important due to the need for organizing and data management of digital information. These classification systems are often created by experts in a certain subject, or technical experts whom "frequently join forces through committees or designated initiatives to design a metadata scheme," (Greenberg, p.22, 2005). While there are hundreds of different metadata schemes, there is no comprehensive list of those that are currently in use (Greenberg, 2005). Therefore, this paper will give brief overviews of the metadata schemes that are most widely used, both for digital information and for physical information, including the MODS scheme, Dublin Core, TEI, and METS schemes.

MARC, or Machine-Readable Cataloging, was created by a Library of Congress initiative nearly 40 years ago to provide a mechanism by which computers exchange, use, and interpret bibliographic information. Many library catalogs today use MARC 21, the more recent version of the standard MARC cataloging system. MARC records have three elements; the record structure, the content designation (which identifies the semantics of the data elements, usually using numerical codes), and the data content within the record. MARC 21 can carry bibliographic information for many different types of data. This includes maps, books, computer files, or music. However, this scheme tends to contain more information than more user friendly schemes, and the coding process tends to be difficult (Network Development & MARC Standards Office, 2006).

In response to the difficulty of using and maintaining information with MARC schemes, the Library of Congress created a new description scheme known as Metadata Object Description Schema (MODS). MODS were created in 2002 in an attempt to create a more user-friendly version of MARC 21. XML based, this scheme uses language tags instead of numerical tags, and multiple data elements in MARC form can often be condensed to one element (Network Development & MARC Standards Office, 2010). MODS is "a derivative of MARC 21 and intended to either carry selected data from existing MARC 21 records or enable the creation of original resource description records." . While MODS can stand alone, because of its flexibility in formatting, it can also complement other metadata schemes such as METS (NISO, 2004).

While MARC 21 and MODS schemas were created to be used to describe a variety of different types of information, the Dublin Core Metadata Element Set, created in 1995, was

developed to describe Web-based documents. While the scheme can, in fact, support other types of data, it is widely used for cataloging and organizing information from the World Wide Web;

> The original objective of the Dublin Core was to define a set of elements that could be used by authors to describe their own Web resources. Faced with a proliferation of electronic resources and the inability of the library profession to catalog all these resources, the goal was to define a few elements and some simple rules that could be applied by noncatalogers. The original 13 core elements were later increased to 15:*Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage,* and *Rights* (NISO, p. 3, 2004).

While some disagreement remains on how simplistic the Dublin Core scheme should be, it is widely known as the most user-friendly metadata scheme.

The Text Encoding Initiative (TEI) scheme was created through an international project in an effort to make guidelines regarding the marking of electronic texts such as novels, plays, or poetry. TEI specifies how to encode text, and how to specify a header which consists of the basic metadata about the work. TEI encoded texts are assumed to be electronic versions of printed texts, and therefore have their own metadata scheme. Within the TEI encoded texts, bibliographic information can be held for both the printed and digital texts. TEI is easily mapped to and from MARC, making the transference of information easier. Libraries tend to use TEI headers when they have electronic collections, and often use a combination of TEI and MARC headers in their catalogs (NISO, 2004).

The final metadata scheme that is often used today is the Metadata Encoding and Transmission Standard (METS), which filled the need for a standard data structure for "describing complex digital library objects" (NISO, p.4, 2004). METS creates XML documents

to organize the majority of administrative metadata and structural metadata in libraries. This includes the names and locations of files and the management of digital objects. The use of METS is widespread, given that it covers an extensive amount of information that was either not covered or not formally covered in previous metadata schemes (NISO, 2004).

Conclusion

Metadata has many practical applications, extending from schemes that were designed for organization and preservation. By looking at the definition of metadata from a librarian's perspective, the applications of metadata as an organization system become more apparent. The uses of metadata are more than just organizational, serving to advance research as well as archive and preserve older documents for future researchers. Metadata can be presented as structural, descriptive and administrative, each type providing different types of information about the document in question. There are many schemes developed for presenting metadata, each one providing the same information in different displays depending on the semantics, syntax and content. Metadata's uses extend beyond the many uses in libraries, also providing services for museums, digital libraries and archiving libraries.

Bibliography

1. Dutta, B. "Cataloging Web Documents using Dublin Core, Marc21." *Natural Sciences Library*. Indian Statistical Institute, n.d. Web. 15 Nov 2012.
2. Greenburg, J. "Understanding Metadata and Metadata Schemes." *Cataloging and Classification Quarterly*. 2005: 17-36. Web. 20 Nov. 2012.
3. Harpring, Patricia. "Introduction to Metadata: Crosswalk." *Getty Research Institute*. Getty Research Institute, 9 2009. Web. 16 Nov 2012. <http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html>.
4. National Information Standards Organization. "Understanding Metadata." (2004): 1-16. *NISO*. Web. 30 Nov 2012. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>.
5. Network Development & MARC Standards Office. Library of Congress. *Marc 21 Format for Bibliographic Data*. Washington: Cataloging Distribution Service, 2006. Web. <http://www.loc.gov/marc/bibliographic/bdintro.html>.
6. Network Development & MARC Standards Office. Library of Congress. *Mods: Uses and features*. Washington: Cataloging Distribution Service, 2010. Web. <http://www.loc.gov/standards/mods/mods-overview.html>.
7. Robertson, James. "Metadata fundamentals for intranets and websites." *Step Two Designs*. Step Two Designs, 19 2008. Web. 20 Nov 2012. <http://www.steptwo.com.au/papers/kmc_metadata/index.html>.