Metadata Schemes Points of Comparison
©2006 Jeffrey Beall

This article lists the chief points of comparison among the different metadata schemes available. Before implementing a metadata scheme, digital libraries must decide on which one to use. Knowing the chief points of comparison among the schemes available can help with this process. Some digital libraries have chosen to create new schemes instead of implementing an existing one, when existing schemes may have been adequate for their needs. Knowing the points of comparison among metadata schemes is also valuable when an institution is evaluating the effectiveness of a scheme already in use.

Granularity and Formats of Description

Metadata schemes differ in the amount of specificity they provide for and their ability to describe data in different formats. For example, some schemes provide a way of differentiating among different types of authors, yet others do not. Different types of authors include personal authors, corporate authors, and conference authors. Here the *specificity* is also often referred to as *granularity*. [1] Schemes also differ in their ability to describe data that comes in different formats. For example, some schemes may only be designed to describe data in electronic form, and others can describe data in any form.

Level of Connection to Content Standards

Some schemes, like MARC, are closely connected to content standards. MARC is often closely associated with the *Anglo-American Cataloguing Rules* and with the *Library of Congress Subject Headings*. Other schemes are much less connected to content standards, so selecting such a scheme may also involve the additional task of selecting content standards. On the other hand, selecting a scheme with a strong connection to a particular content standard means having also to adopt the content standards, ontologies, etc. that are associated with it.

Also, schemes may differ in their ability to encode different types of data, such as non-Roman scripts, Unicode, etc., but this ability may also depend on the computer system being used to encode the data. Content standard selection is important because it can affect the ability to crosswalk data from one database into another.

Availability of Searching Systems

Metadata systems include software or applications that set up a search interface for metadata. Integrated library systems (ILSs) are an example of a system that searches MARC metadata. One problem with some of the less popular metadata schemes is that there is a lack of systems available to fully exploit the metadata and create a search platform for it.

Another aspect of metadata systems is metadata creation. Potential implementers should determine whether the scheme in question has systems available for metadata creation by humans or computers. An example of this is integrated library systems that have the functionality to create MARC records. Similarly, systems' differ in their ability to store and manipulate data created in a particular scheme.

The next few years will likely see a greater development of digital library management systems (DLMSs) [2] that will differ in their ability to accommodate different metadata schemes. The process of selecting a particular scheme will need to take into account the availability of systems for a given scheme.

Level of Community Specificity

Some metadata schemes are created for the specific needs of an individual community. For example, the MPEG-7 scheme is designed for multimedia. The ONIX scheme is designed for the book trade industry. Other schemes are general in design, and can accommodate metadata from most fields of study. The desire for community specificity has led to an abundance of metadata schemes.

Interoperability

This characteristic describes several things. First, it describes how well-suited a scheme is for crosswalking data into other schemes. More practically, it involves whether those mappings have been made and are available. Most of the more popular schemes have had crosswalks to other schemes developed. For example, there is a crosswalk from Dublin Core to EAD.

Interoperability also includes metadata harvesting. A scheme with high interoperability will be one that enables the harvesting and meta-searching of metadata encoded in it. To some degree, interoperability is related to a scheme's popularity. The more popular and widely used a scheme is, the more likely it is to have crosswalks to other schemes and harvesting standards.

Proven Success, Reputation, Popularity

Success and popularity of a scheme often weigh heavily for users deciding whether or not to adopt a particular scheme. Users will likely prefer to select a scheme that has successfully left beta testing and has had at least several documented, successful implementations.

Amount of Training Required

Those selecting a scheme will need to take into account the amount of training individuals will need to become proficient in encoding metadata in the scheme. For schemes that are closely connected to content standards, this training will also need to take into account the amount of training needed to gain proficiency in those standards. There is likely a positive correlation between the amount of training needed to master a scheme and the richness of description it provides.

Viability of the Organization behind the Scheme

The stability and vibrancy of the organizations behind metadata schemes are crucial to their success. Potential implementers of a scheme should investigate the organization behind it and make sure that it keeps the scheme current with the latest developments and user needs. A related factor worth investigating is how open the organization is to receiving input and suggestions from implementers. Also, implementers will need to consider the amount and quality of documentation that is available for a particular scheme, as well how up-to-date it is.

Ability of the Scheme to Handle a Particular Metadata Function

Metadata serves different purposes. Some include discovery, rights management, and preservation data. But not all schemes are able to serve all of the various functions. Before implementing a scheme, users need to determine exactly what functions they want their metadata to serve, and then they should select a scheme that adequately handles these functions.

Adaptability of the Scheme to Local Needs

This relates to community specificity but is different in that some metadata schemes can be changed at the local level, such as by adding certain new fields or tags. Sometimes a modified scheme is also called a particular "flavor" of a scheme. For example, the Collaborative Digitization Program has created the Western States Dublin Core, which is a customized implementation of Dublin Core. Schemes that are more adaptable will have mechanisms for extensibility of the data elements so that they can be extended to meet local needs.

Scalability

Scalability refers to how large a database of metadata the scheme and its retrieval system can handle successfully. For example, a scheme with only a few elements of description is not as scalable as a system with many elements because when you have millions of records using a "few-element" scheme, it becomes harder to generate precise search results. In general, the richer the description a scheme provides for, the more scalable it is.

Surrogacy

This relates only to digital objects and describes whether the metadata is embedded in the object it describes or exists separately from it in a searchable database. Howarth first describes metadata that does not exist as a surrogate for the object it describes:

> In general, a distinction can be made between simple format metadata – such as that represented in the syntax of a mark-up language (e.g., XML; HTML; SGML), and embedded within the structure of the digital object – and structured rich format metadata. For the former, Web crawlers or "bots" can harvest the specified metatags (e.g., <Title>) to extract particular values … [3]

Of course, some schemes can have the metadata exist within the data it describes and also as a surrogate separate from it.

Conclusion

As the number of metadata schemes continues to grow, digital libraries will need clear points of comparison for selecting and evaluating from among the schemes available. The points listed here can serve as the basis for making an implementation decision or for evaluating an existing metadata scheme implementation.

References

[1] See also: Choosing a Metadata Standard for Resource Discovery: http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/html/

[2] Ioannidis, Yannis, 2005, "Digital Libraries at a Crossroads." *International Journal on Digital Libraries 5* no. 4:255-265.

[3] Howarth, Lynne C., 2004, "Metadata Schemas for Subject Gateways." *International Cataloguing and Bibliographic Control 33* no. 1:8-12.