

Metadata:

An Overview

by Kevin Butterfield

In order to be organized, information needs to be described. The language used to describe information is a bibliographic language, a special-purpose language designed and applied according to a special set of rules. The vocabulary of a bibliographic language consists of expressions used to name the values of three variables: entities, attributes and relationships. In languages used to describe the physical and production attributes of documents, these vocabulary terms are referred to as data elements or metadata.

Metadata is commonly defined as data about data. It may be further defined as an expansion of bibliographic cataloging practices in a digital environment. In particular communities and contexts, however, the word is used with much narrower definitions. For example, it may mean only cataloging, or only data about digital resources, or only information structured to be understood by computers. However, none of these limits is intrinsic to metadata. It is important to set the definition of the term for any particular discussion to avoid misunderstanding. It is also useful to remember when discussing metadata to define whether the discussion includes tag or field sets, the syntax for entering data into those sets or both. While metadata schemes can be very strict about the use of elements, their order and attributes — author, creator, title, etc. — they can also be very vague or provide no guidance as to how data should be entered into those fields or elements. This is similar to trying to describe an item using Machine Readable Cataloging record (MARC) but not the Anglo-American Cataloging Rules, 2nd ed. (AACR2).

The semantics of the data vary across constituencies. Different disciplines require different types of data. No one group will be able to impose order on them all. The trend is toward multiple communities developing standards and allowing for bridging mechanisms to connect them. However, there is a core set of terms that cuts across each metadata scheme: author, title, publisher, etc. The names of the individual elements may vary — author, creator, etc. — but the concepts remain closely tied.

Describing digital legal objects such as cases, briefs or opinions actually comes easily to this type of scheme. Content such as this is very structured. Each case or opinion follows a similar format. Encoding this type of information flows naturally from the front matter. Tags such as <appellee>, <appellant>, <jurisdiction>, <date type=argued>, <date type=decided>, <judge type=chief>, <judge type=circuit> are suggested by reviewing an opinion. Standardizing tag sets and developing Document Type Definitions for these types of legal publications represent a major step forward.

The use of metadata to discover the existence of resources online is only a starting point. Users often need to know about a resource's history; its ownership; its accuracy; its integrity; its relationship to other sources; its intellectual property rights; its technical features regarding how to run, display or otherwise use it; the terms and conditions under which the material is available; and many other such matters related to usefulness, usability and economics. Metadata is far more than descriptive and it's not necessarily static. For additional metadata can accrue over time as individual resources are used and as their status changes. The Text Encoding Initiative header, for example, includes a Revision Description. This section of the header includes a description of all revisions made to the text as it passes from user to user. The metadata thus exists as a living component of the document.

Functional Groupings of Metadata

In addition to organizing metadata by specific elements, functional areas can group metadata as well. Anne Gilliland-Sweetland, in *Introduction to Metadata: Pathways to Digital Information*, divides metadata into the following five functional areas.

Administrative metadata supports resource management within a collection and is used in managing and administering information resources. This group may include:

- Acquisitions information
- Rights and reproduction tracking
- Documentation of legal-access requirements
- Location information
- Selection criteria for digitization
- Version control

Descriptive metadata facilitates resource discovery and identification and is used to describe or identify information resources. This group may include:

- Cataloging records
- Finding aids
- Specialized indexes
- Hyperlinked relationships between resources
- Annotations by users

This is the most well-known of these functional groups as it bears the most resemblance to the work done in libraries. Efforts such as the Dublin Core and Encoded Archival Description fall into this area.

Preservation metadata facilitates the preservation management of information resources. This group may include:

- Documentation of the physical condition of the resources
- Documentation of the actions taken to preserve physical and/or digital versions of resources — e.g., data refreshing and migration
- Documentation of the technical processes associated with preserving the item
- Specification of the rights management information
- Establishment of the authenticity of digital content
- Recording the chain of custody for a digital object
- Unique identification — both internally and externally in relation to the archive in which it resides

Technical metadata relates to how a system functions or to how metadata behaves. This group may include:

- Hardware and software documentation
- Digitization information, formats, compression ratios, scaling routines
- Tracking of system response times
- Authentication and security data, encryption keys, passwords

The current National Information Standards Organization draft standard, *Data Dictionary: Technical Metadata for Digital Still Images*, attempts to address these issues for image files by defining metadata to, for example, identify the data elements that would be used by applications to control transformations of images against stated metrics for meaningful quality attributes, such as detail, tone, color and size.

Usage metadata tracks the level and type of use of information resources. This group may include:

- Use and user tracking
- Circulation records
- Content reuse and multi-versioning information

In the past, these different groups of metadata existed independently of each other in files and card catalogs. As libraries began automating catalogs and circulation functions, the information moved closer together but still existed independently of the item. Now, with digital objects, it is possible to integrate metadata with the object itself. Either in the form of a header or an Resource Description Framework packet, the metadata can accrue and travel throughout the life of the item it describes.

Why Continue to Develop Metadata?

The various reasons why we continue to work with and develop metadata schemes reflect the same needs libraries have for cataloging and managing items in our collections. Gilliland-Sweetland's outline of the reasons are summarized below.

Increased Accessibility. Rich, consistent metadata can greatly increase effectiveness

of searching, make it possible to search across multiple collections, and create virtual collections from materials that are distributed across several repositories, but only if the descriptive metadata can be mapped across each site. Great efforts have been made to create crosswalks between various metadata schema and MARC in an effort to allow systems to gather and display metadata regardless of format. These crosswalks also allow various constituencies to continue developing and working with metadata in their chosen format. In addition to crosswalks, efforts such as the Open Archives Initiative develops means to register and share metadata across institutions.

Retention of Context. Objects that are a part of a collection may lose connections and associations with other objects, people, etc. Metadata can be used to document and maintain relationships and indicate authenticity, structural integrity, and comprehensiveness of information objects.

Expanding Use. Digitizing resources makes them more accessible — freeing access to the object the users or information seekers from geographic constraints. New communities of users, however, may have needs that differ from the needs of traditional users around whom many existing information services have been designed. The emerging self-help legal market is an example of this. The Information Management Group to the Project for the Future of Equal Justice has put a great deal of effort into designing metadata and portals guiding users who lack formal legal training to the information they seek. This concept becomes more interesting as multiple languages and

cultural perspectives are introduced. As objects become more universally accessible, the need grows for a descriptive method that is adaptable to these varied audiences.

Multi-versioning. Metadata may link together multiple versions and capture the similarities and differences of each version. In the case of images, a high-resolution copy may be kept for preservation purposes and a lower, more easily transferable copy delivered to users. The links between these and the original photograph or item need to be maintained.

Legal Issues. Metadata can be used for tracking rights — Internet Protocol and copyright — for objects and multiple versions, for legal or donor requirements imposed on objects. Digital Object Identifiers, for example, make wide use of metadata in enabling automated copyright management.

Preservation. Metadata enables an object to exist independently of the system currently being used to store it. Technical, descriptive and preservation metadata document how a digital information object was created and maintained, how it behaves, and how it relates to other objects. These data must be preserved and migrated.

The containers used to package descriptions of legal information may vary, but the issues remain the same.

Kevin Butterfield (butterfi@law.uivc.edu) is head of technical services and assistant professor of library administration at Albert E. Jenner Jr. Memorial Law Library of the University of Illinois at Urbana-Champaign.