Setting the Stage

Anne J. Gilliland

Metadata, literally "data about data," has become a widely used yet still frequently underspecified term that is understood in different ways by the diverse professional communities that design, create, describe, preserve, and use information systems and resources. It is a construct that has been around for as long as humans have been organizing information, albeit transparently in many cases, and today we create and interact with it in increasingly digital ways. For the past hundred years at least, the creation and management of metadata has primarily been the responsibility of information professionals engaged in cataloging, classification, and indexing; but as information resources are increasingly put online by the general public, metadata considerations are no longer solely the province of information professionals. Although *metadata* is arguably a much less familiar term among creators and consumers of networked digital content who are not information professionals per se, these same individuals are increasingly adept at creating, exploiting, and assessing user-contributed metadata such as Web page title tags, folksonomies, and social bookmarks. Schoolchildren and college students are taught in information literacy programs to look for metadata such as provenance and date information in order to ascertain the authoritativeness of information that they retrieve on the Web. Thus it has become more important than ever that not only information professionals but also other creators and users of digital content understand the critical roles of different types of metadata in ensuring accessible, authoritative, interoperable, scaleable, and preservable cultural heritage information and record-keeping systems.

Until the mid-1990s, *metadata* was a term used primarily by communities involved with the management and interoperability of geospatial data and with data management and systems design and maintenance in general. For these communities, *metadata* referred to a suite of industry or disciplinary standards as well as additional internal and external documentation and other data necessary for the identification, representation, interoperability, technical management, performance, and use of data contained in an information system.

Perhaps a more useful, "big picture" way of thinking about metadata is as the sum total of what one can say about any *information object* at any level of aggregation.[1] In this context, an information object is anything that can be addressed and manipulated as a discrete entity by a human being or an information system. The object may comprise a single item, it may be an aggregate of many items, or it may be the entire database or record-keeping system. Indeed, in any given instance one can expect to find metadata relevant to any information object existing simultaneously at the item, aggregation, and system levels.

In general, all information objects, regardless of the physical or intellectual form they take, have three features—content, context, and structure—all of which can and should be reflected through metadata.

- *Content* relates to what the object contains or is about and is *intrinsic* to an information object.
- *Context* indicates the who, what, why, where, and how aspects associated with the object's creation and is *extrinsic* to an information object.
- *Structure* relates to the formal set of associations within or among individual information objects and can be *intrinsic* or *extrinsic* or both.

Cultural heritage information professionals such as museum registrars, library catalogers, and archival processors often apply the term *metadata* to the value-added information that they create to arrange, describe, track, and otherwise enhance access to information objects and the physical collections related to those objects. Such metadata is frequently governed by community-developed and community-fostered standards and best practices in order to ensure quality, consistency, and interoperability. The following Typology of Data Standards organizes these standards into categories and provides examples of each. Markup languages such as HTML and XML provide a standardized way to structure and express these standards for machine processing, publication, and implementation.

Library metadata development has been first and foremost about providing intellectual and physical access to collection materials. *Library metadata* includes indexes, abstracts, and bibliographic records created according to cataloging rules (data content standards) such as the *Anglo-*

[1] An information object is a digital item or group of items, regardless of type or format, that can be addressed or manipulated as a single object by a computer. This concept can be confusing in that it can be used to refer both to digital "surrogates" of original objects or items (e.g., digitized images of works of art or material culture, a PDF of an entire book) and to descriptive records relating to objects and/or collections (e.g., catalog records or finding aids).

Table 1. **A Typology of Data Standards**

Type	Examples
Data *structure* standards (metadata element sets, schemas). These are "categories" or "containers" of data that make up a record or other information object.	The set of MARC (Machine-Readable Cataloging format) fields, Encoded Archival Description (EAD), Dublin Core Metadata Element Set (DCMES), Categories for the Description of Works of Art (CDWA), VRA Core Categories
Data *value* standards (controlled vocabularies, thesauri, controlled lists). These are the terms, names, and other values that are used to populate data structure standards or metadata element sets.	Library of Congress Subject Headings (LCSH), Library of Congress Name Authority File (LCNAF), LC Thesaurus for Graphic Materials (TGM), Medical Subject Headings (MeSH), Art & Architecture Thesaurus (AAT), Union List of Artist Names (ULAN), Getty Thesaurus of Geographic Names (TGN), ICONCLASS
Data *content* standards (cataloging rules and codes). These are guidelines for the format and syntax of the data values that are used to populate metadata elements.	Anglo-American Cataloguing Rules (AACR), Resource Description and Access (RDA), International Standard Bibliographic Description (ISBD), Cataloging Cultural Objects (CCO), Describing Archives: A Content Standard (DACS)
Data *format/technical interchange* standards (metadata standards expressed in machine-readable form). This type of standard is often a manifestation of a particular data structure standard (type 1 above), encoded or marked up for machine processing.	MARC21, MARCXML, EAD XML DTD, METS, MODS, CDWA Lite XML schema, Simple Dublin Core XML schema, Qualified Dublin Core XML schema, VRA Core 4.0 XML schema

Note: This table is based on the typology of data standards articulated by Karim Boughida, "CDWA Lite for Cataloging Cultural Objects (CCO): A New XML Schema for the Cultural Heritage Community," in *Humanities, Computers and Cultural Heritage: Proceedings of the XVI International Conference of the Association for History and Computing:* 14–17 (September 2005) (Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005). Available at http://www.knaw.nl/publicaties/pdf/20051064.pdf.

American Cataloguing Rules (AACR) and data structure standards such as the MARC (*Machine-Readable Cataloging*) format, as well as data value standards such as the *Library of Congress Subject Headings* (LCSH) or the *Art & Architecture Thesaurus* (AAT). Such bibliographic metadata has been systematically and cooperatively created and shared since the 1960s and made available to repositories and users through automated systems such as bibliographic utilities, online public access catalogs (OPACs), and commercially available databases. Today this type of metadata is created not only by humans but also in automated ways through such means as metadata mining, metadata harvesting, and Web crawling. Automation of metadata will inevitably continue to expand with the development of the Resource Description Framework (RDF) and the Semantic Web, which are discussed later in this book.

A large component of archival and museum metadata creation activities has traditionally been focused on context. Elucidating and preserving context is what assists with identifying and preserving the evidential value of records and artifacts in and over time; it is what facilitates the authentication of those objects, and it is what assists researchers with their analysis and interpretation. *Archival and manuscript metadata* (more commonly referred to as *archival description*) includes accession

records, finding aids, and catalog records. Archival data structure standards that have been developed in the past three decades include the MARC Archival and Manuscripts Control (AMC) format, published by the Library of Congress in 1984 (now integrated into the MARC21 format for bibliographic description); the General International Standard Archival Description (ISAD (G)), published by the International Council on Archives in 1994; Encoded Archival Description (EAD), adopted as a standard by the Society of American Archivists (SAA) in 1999, and its companion data content standard, *Describing Archives: A Content Standard* (DACS), first published in 2004. The *Metadata Encoding and Transmission Standard* (METS), developed by the Digital Library Federation and maintained by the Library of Congress, is increasingly being used for encoding descriptive, administrative, and structural metadata and digital surrogates at the item level for objects such as digitized photographs, maps, and correspondence from the collections described by finding aids and other collection- or group-level metadata records. While archival metadata was primarily only available locally at individual repositories until the late 1990s, it is now distributed online through resources such as OCLC (Online Computer Library Center),[2] Archives USA,[3] and EAD-based resources such as the Online Archive of California and the Library of Congress's American Memory Project.[4]

Consensus and collaboration have been slower to build in the museum community, where the benefits of standardization of description such as shared cataloging and exchange of descriptive data were less readily apparent until relatively recently. Since the late 1990s, tools such as *Categories for the Description of Works of Art* (CDWA), Spectrum, the CIDOC Conceptual Reference Model, *Cataloging Cultural Objects* (CCO), and the CDWA Lite XML schema have begun to be considered and implemented by museums. Initiatives such as Museums and the Online Archive of California (MOAC)[5] have examined the applicability and extensibility of descriptive standards developed by archives and libraries such as EAD and METS to museum holdings in order to address the integration of cultural information across repository types, as well as the educational needs of users visiting online museum resources.

Although it would seem to be a desirable goal to integrate materials of different types that are related by provenance or subject but distributed across museum, archives, and library repositories, initiatives such as MOAC have met with only limited success. As MOAC and the

---

[2] http://www.oclc.org/.

[3] http://archives.chadwyck.com/.

[4] http://www.oac.cdlib.org/ and http://memory.loc.gov/ammem/index.html.

[5] http://www.bampfa.berkeley.edu/moac/.

mid-1980s development of the now-defunct MARC AMC format have demonstrated, the distinctiveness of the various professional and object-based approaches (e.g., widely differing notions of provenance and collectivity as well as of structure) and the different institutional cultures have left many professionals feeling that their practices and needs have been shoehorned into structures that were developed by another community with quite different practices and users. As enunciated in Principle 6 of "Practical Principles for Metadata Creation and Maintenance" (p. 72), there is no single metadata standard that is adequate for describing all types of collections and materials; selection of the most appropriate suite of metadata standards and tools, and creation of clean, consistent metadata according to those standards, not only will enable good descriptions of specific collection materials but also will make it possible to map metadata created according to different community-specific standards, thus furthering the goal of interoperability discussed in subsequent chapters of this book.

An emphasis on the structure of information objects in metadata development by these communities has perhaps been less overt. However, structure has always been important in information organization and representation, even before computerization. Documentary and publication forms have evolved into industry standards and societal norms and have become an almost transparent information management tool. For example, when users access a birth certificate they can predict its likely structure and content. When academics use a scholarly monograph, they understand intuitively that it will be organized with a table of contents, chapter headings, and an index. Archivists use the physical structure of their finding aids to provide visual cues to researchers about the structural relationships between different parts of a record series or manuscript collection. Archival description also exploits the hierarchical arrangement of records according to the bureaucratic hierarchies and business practices of the creators of those records. However, in recent years there has been increasing criticism that while valuable for retaining context and original order, collection-level, hierarchical metadata as exemplified in archival finding aids privileges the scholarly user of the archive (and those who are familiar with the structure and function of archival finding aids) while leaving the nonexpert user baffled, as well as unnecessarily perpetuating a paper-based descriptive paradigm.[6] In the online world, multiple descriptive relationships between objects can be supported simultaneously, and some of these may more effectively support new types of users and uses in

[6] Anne J. Gilliland-Swetland, "Popularizing the Finding Aid: Exploiting EAD to Enhance Online Browsing and Retrieval in Archival Information Systems by Diverse User Groups," *Journal of Internet Cataloging* 4, nos. 3–4 (2001): 199–225.

an environment that is not mediated by a reference archivist. Archives and other collecting institutions are beginning to explore methods of description that exploit item-level metadata for digitized objects so that users can search for specific items, navigate through a collection "bottom-up" as well as "top-down," and collate related collection materials through lateral searching across collections and repositories.

The role of structure has been growing as computer-processing capabilities become increasingly powerful and sophisticated. Information communities are aware that the more highly structured an information object is, the more that structure can be exploited for searching, manipulation, and interrelating with other information objects. Capturing, documenting, and enforcing that structure, however, can only occur if supported by specific types of metadata. In short, in an environment where a user can gain unmediated access to information objects over a network, metadata

- certifies the authenticity and degree of completeness of the content;
- establishes and documents the context of the content;
- identifies and exploits the structural relationships that exist within and between information objects;
- provides a range of intellectual access points for an increasingly diverse range of users; and
- provides some of the information that an information professional might have provided in a traditional, in-person reference or research setting.

But there is more to metadata than description and resource discovery. A more inclusive conceptualization of metadata is needed as we consider the range of activities that may be incorporated into digital information systems. Repositories also create metadata relating to the administration, accessioning, preservation, and use of collections. Acquisition records, exhibition catalogs, licensing agreements, and educational metadata are all examples of these other kinds of metadata and data. Integrated information resources such as virtual museums, digital libraries, and archival information systems include digital versions of actual collection content (sometimes referred to as *digital surrogates*), as well as descriptions of that content (i.e., descriptive metadata, in a variety of formats). Incorporating other types of metadata into such resources reaffirms the importance of metadata in administering collections and maintaining their intellectual integrity both in and over time. Paul Conway alludes to this capability of metadata when he discusses the impact of digitization on preservation:

> The digital world transforms traditional preservation concepts from protecting the physical integrity of the object to specifying the creation and maintenance of the object whose intellectual integrity is its primary characteristic.[7]

When applied outside the original repository, the term *metadata* acquires an even broader scope. An Internet resource provider might use *metadata* to refer to information that is encoded in HTML META tags for the purposes of making a Web site easier to find. Individuals who are digitizing images might think of metadata as the information they enter into a header field for the digital file to record information about the image file, the imaging process, and image rights. A social science data archivist might use the term to refer to the systems and research documentation necessary to run and interpret a magnetic tape containing raw research data. An electronic records archivist might use the term to refer to all the contextual, processing, preservation, and use information needed to identify and document the scope, authenticity, and integrity of an active or archival record in an electronic record-keeping or archival preservation system. Metadata is crucial in personal information management and for ensuring effective information retrieval and accountability in record keeping—something that is becoming increasingly important with the rise of electronic commerce and the use of digital content and tools by governments. In all these diverse interpretations, metadata not only identifies and describes an information object; it also documents how that object behaves, its function and use, its relationship to other information objects, and how it should be and has been managed over time.

As this discussion suggests, theory and practices vary considerably due to the differing professional and cultural missions of museums, archives, libraries, and other information and record-keeping communities. Information professionals have a bewildering array of metadata standards and approaches from which to choose. Many highly detailed metadata standards have been developed by individual communities (e.g., MARC, EAD, the Australian Recordkeeping Metadata Schema, RKMS, and some of the standards for Geographic Information Systems) that attempt to articulate their mission-specific differences as well as to facilitate mapping between common data elements. If used appropriately and to their fullest extent, these standards have the potential to create extremely rich metadata that would provide detailed documentation of record-keeping creation and

---

[7] Paul Conway, *Preservation in the Digital World* (Washington, DC: Commission on Preservation and Access, 1996). http://www.clir.org/pubs/reports/conway2/index.html.

[8] Sue McKemmish, Glenda Acland, Nigel Ward, and Barbara Reed, "Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema," *Archivaria* 48 (Fall 1999): 3–37.

use in situations in which such activities may be challenged or audited for their comprehensiveness and accuracy.[8] Creation and ongoing maintenance of such metadata, however, is complex, time consuming, and resource intensive and may only be justifiable when there is a legal mandate or other risk management incentive or when it is envisaged that the content and metadata may be reused or exploited in previously unanticipated ways, such as in digital asset management systems. By contrast, the Dublin Core Metadata Element Set (DCMES) identifies a relatively small, generic set of metadata elements that can be used by any community, expert or nonexpert, to describe and search across a wide variety of information resources on the World Wide Web. Such metadata standards are necessary to ensure that different kinds of descriptive metadata are able to interoperate with one other and with metadata from nonbibliographic systems of the kind that the data management communities and information creators are generating. Relatively lean metadata records such as those created using the DCMES have the advantage of being cheaper to create and maintain, but they may need to be augmented by other types of metadata in order to address the needs of specific user communities and to adequately describe particular types of collection materials.[9]

Another form of metadata that has recently begun to appear is user created; user-created metadata has been gathering momentum in a variety of venues on the Web. Just as many members of the general public have participated in the development of Web content, whether through personal Web pages or by uploading photos onto Flickr or videos onto YouTube, they have also increasingly been getting into the business of creating, sharing, and copying metadata (albeit often unknowingly). Folksonomies that are created using specialized tagging tools in various Web-based communities in order to identify, retrieve, categorize, and promote Web content and the sharing of bookmarks through the practice of social bookmarking are examples of the burgeoning user-created metadata on the Web. Among the advantages of these approaches is that individual Web communities such as affinity groups or hobbyists may be able to create metadata that addresses their specific needs and vocabularies in ways that information professionals who apply metadata standards designed to cater to a wide range of audiences cannot. User-generated metadata is also a comparatively inexpensive way to augment existing metadata, with the cost and the sense of ownership shared among more parties than just those who create information repositories. The disadvantages of user-generated metadata relate to quality control (or lack thereof) and idiosyncrasies

[9] See Roy Tennant, "Metadata's Bitter Harvest," *Library Journal*, August 15, 2004, available at http://www.libraryjournal.com/article/CA434443.html; and the Digital Library Federation's Multiple Metadata Formats page at http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/MultipleMetadataFormats.

that can impede the trustworthiness of both metadata and the resource it describes and negatively affect interoperability between metadata and the resources it is intended to describe. Issues of interoperability are discussed in some detail in the third chapter of this book.

Categorizing Metadata

All these perspectives on metadata should be considered in the development of networked digital information systems, but they lead to a very broad and often confusing conception. To understand this conception better, it is helpful to separate metadata into distinct categories—administrative, descriptive, preservation, use, and technical metadata—that reflect key aspects of metadata functionality. Table 2 defines each of these metadata categories and gives examples of common functions that each might perform in a digital information system.

Table 2. **Different Types of Metadata and Their Functions**

Type	Definition	Examples
Administrative	Metadata used in managing and administering collections and information resources	• Acquisition information • Rights and reproduction tracking • Documentation of legal access requirements • Location information • Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	• Cataloging records • Finding aids • Differentiations between versions • Specialized indexes • Curatorial information • Hyperlinked relationships between resources • Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	• Documentation of physical condition of resources • Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration • Documentation of any changes occurring during digitization or preservation
Technical	Metadata related to how a system functions or metadata behaves	• Hardware and software documentation • Technical digitization information, e.g., formats, compression ratios, scaling routines • Tracking of system response times • Authentication and security data, e.g., encryption keys, passwords
Use	Metadata related to the level and type of use of collections and information resources	• Circulation records • Physical and digital exhibition records • Use and user tracking • Content reuse and multiversioning information • Search logs • Rights metadata

Table 3. **Attributes and Characteristics of Metadata**

Attribute	Characteristics	Examples
Source of metadata	Internal metadata generated by the creating agent for an information object at the time when it is first created or digitized	• File names and header information • Directory structures • File format and compression scheme
	Metadata intrinsic to an item or work	• A title or other inscription added to an art work by its creator • A title or subtitle on the title page of a manuscript or printed book
	External metadata relating to an original item or information object, that is created later, often by someone other than the original creator	• URLs and other digital statements of provenance • "Tracked changes" • Registrarial and cataloging records • Rights and other legal information
Method of metadata creation	Automatic metadata generated by a computer	• Keyword indexes • User transaction logs • Audit trails
	Manual metadata created by humans	• Descriptive metadata such as catalog records, finding aids, and specialized indexes
Nature of metadata	Nonexpert metadata created by persons who are neither subject specialists nor information professionals, e.g., the original creator of the information object or a folksonomist	• META tags created for a personal Web page • Personal filing systems • Folksonomies
	Expert metadata created by subject specialists and/or information professionals, often not the original creator of the information object	• Specialized subject headings • MARC records • Archival finding aids • Catalog entries for museum objects • Ad hoc metadata created by subject experts, e.g., notations by scholars or researchers
Status	Static metadata that does not or should not change once it has been created	• Technical information such as the date(s) of creation and modification of an information object, how it was created, file size
	Dynamic metadata that may change with use, manipulation, or preservation of an information object	• Directory structure • User transaction logs
	Long-term metadata necessary to ensure that the information object continues to be accessible and usable	• Technical format and processing information • Rights information • Preservation management documentation
	Short-term metadata, mainly of a transactional nature	• Interim location information
Structure	Structured metadata that conforms to a predictable standardized or proprietary structure	• MARC • TEI • EAD • CDWA Lite • Local database formats
	Unstructured metadata that does not conform to a predictable structure	• Unstructured note fields and other free-text annotations

Attribute	Characteristics	Examples
Semantics	Controlled metadata that conforms to a standardized vocabulary or authority form, and that follows standard content (i.e., cataloging) rules	• LCSH, LCNAF, AAT, ULAN, TGM, TGN • AACR (RDA), DACS, CCO
	Uncontrolled metadata that does not conform to any standardized vocabulary or authority form	• Free-text notes • HTML META tags and other user-created tags
Level	Collection-level metadata relating to collections of original items and/or information objects	• Collection- or group-level record, e.g., a MARC record for a group or collection of items; a finding aid for an intact archival collection • Specialized index
	Item-level metadata relating to individual items and/or information objects, often contained within collections	• Catalog records for individual bibliographic items or unique cultural objects • Transcribed image captions and dates • "Tombstone" information for works of art and material culture • Format information

In addition to its different types and functions, metadata exhibits many different characteristics. Table 3 presents some key characteristics of metadata, with examples.

Metadata creation and management have become a complex mix of manual and automatic processes and layers created by many different functions and individuals at different points during the life cycle of an information object. One emergent area is metadata management, the aim of which is to ensure that the metadata we rely on to validate Web resources is itself trustworthy and that the large volume of metadata that potentially can accumulate throughout the life of a Web resource is subject to a summarization and disposition regime.[10]

Figure 1 illustrates the different phases through which information objects typically move during their life cycles in today's digital environment.[11] As they move through each phase in their life cycles, information objects acquire layers of metadata that can be associated with them in several ways. Different types of metadata can become associated with an information object by a variety of processes, both human and

[10] See Anne J. Gilliland, Nadav Rouche, Joanne Evans, and Lori Lindberg, "Towards a Twenty-first Century Metadata Infrastructure Supporting the Creation, Preservation and Use of Trustworthy Records: Developing the InterPARES2 Metadata Schema Registry,"*Archival Science* 5, no. 1 (March 2005): 43–78.

[11] Modified from Information Life Cycle, *Social Aspects of Digital Libraries: A Report of the UCLA-NSF Social Aspects of Digital Libraries Workshop* (Los Angeles, CA: Graduate School of Education and Information Studies, November 1996), p. 7.

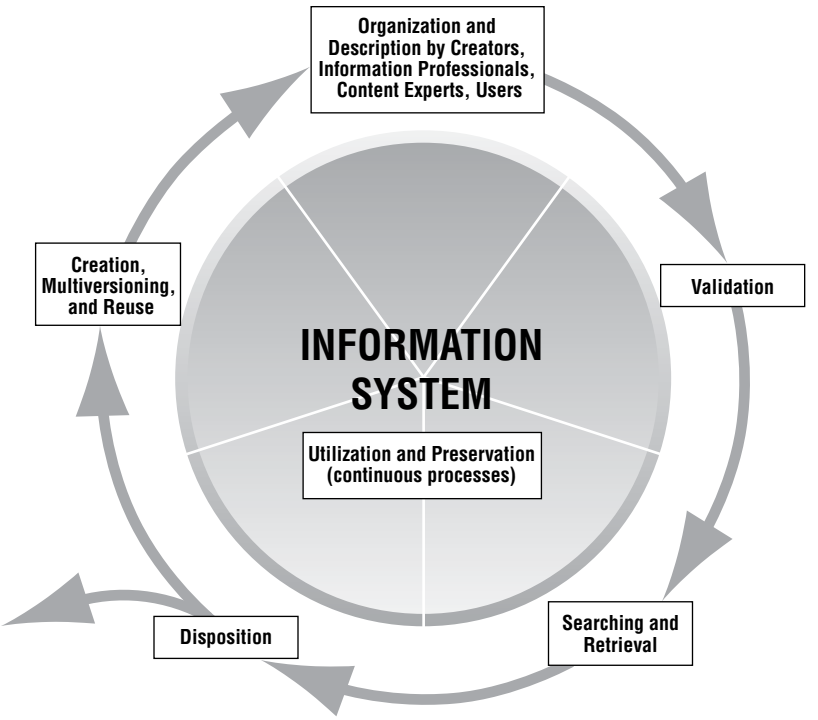

Figure 1. **The Life Cycle of an Information Object**

automated. These layers of accrued metadata can be contained within the same "envelope" as the information object—for example, in the form of header information for an image file or through some form of metadata bundling, for example via METS, which packages structural, descriptive, administrative, and other metadata with an information object or digital surrogate and indicates the types of relationships among the various parts of complex information objects (e.g., a digital surrogate consisting of a series of images representing the pages in a book or in an album of illustrations, or the constituent parts of a decorative arts object such as a tea service). Metadata can also be attached to the information object through bidirectional pointers or hyperlinks, while the relationships between metadata and information objects, and between different aspects of metadata, can be documented by registering them with a metadata registry. However, in any instance in which it is critical that metadata and content coexist, it is highly recommended that the metadata become an integral part of the information object, that is, that it be "embedded" in the object and not stored or linked elsewhere.

As systems designers increasingly respond to the need to incorporate and manage metadata in information systems and to address how to ensure the ongoing viability of both information objects and their associated metadata forward through time, many additional mechanisms for associating metadata with information objects are likely to become available. Metadata registries and schema record-keeping systems are also more likely to develop as it becomes increasingly necessary to document schema evolution and to alert implementers to version changes.[12]

Primary Functions of Metadata

- *Creation, multiversioning, reuse, and recontextualization of information objects.* Objects enter a digital information system by being created digitally or by being converted into digital format. Multiple versions of the same object may be created for preservation, research, exhibit, dissemination, or even product-development purposes. Some administrative and descriptive metadata may and indeed should be included by the creator or digitizer, especially if reuse is envisaged, such as in a digital asset management (DAM) system.
- *Organization and description.* A primary function of metadata is the description and ordering of original objects or items in a repository or collection, as well as of the information objects relating to the originals. Information objects are automatically or manually organized into the structure of the digital information system and may include descriptions generated by the original creator. Additional metadata may be created by information professionals through registration, cataloging, and indexing processes or by others via folksonomies and other forms of user-contributed metadata.
- *Validation.* Users scrutinize metadata and other aspects of retrieved resources in order to ascertain the authoritativeness and trustworthiness of those resources.
- *Searching and retrieval.* Good descriptive metadata is essential to users' ability to find and retrieve relevant metadata and information objects. Locally stored as well as virtually distributed information objects are subject to search and retrieval by users, and information systems create and maintain metadata that tracks retrieval algorithms, user transactions, and system effectiveness in storage and retrieval.

---

[12] See Gilliland et al., "Towards a Twenty-first Century Metadata Infrastructure."

- *Utilization and preservation.* In the digital realm, information objects may be subject to many different kinds of uses throughout their lives, during which processes they may also be reproduced and modified. Metadata related to user annotations, rights tracking, and version control may be created. Digital objects, especially those that are born digital, also need to be subject to a continuous preservation regime and undergo processes such as refreshing, migration, and integrity checking to ensure their continued availability and to document any changes that might have occurred to the information object during preservation processes.
- *Disposition.* Metadata is a key component in documenting the disposition (e.g., accessioning, deaccessioning) of original objects and items in a repository, as well as of the information objects relating to those originals. Information objects that are inactive or no longer necessary may be discarded.

Some Little-Known Facts about Metadata

- *Metadata does not have to be digital.* Cultural heritage and information professionals have been creating metadata for as long as they have been managing collections. Increasingly, such metadata is being incorporated into digital information systems, but metadata can also be recorded in analog formats such as card catalogs, vertical files, and file labels.
- *Metadata relates to more than the description of an object.* While museum, archive, and library professionals may be most familiar with the term in association with description or cataloging, metadata can also indicate the context, management, processing, preservation, and use of the resources being described.
- *Metadata can come from a variety of sources.* Metadata can be supplied by a human (by the creator of the digital file, by an information professional, and/or by an expert or nonexpert user). It can also be generated automatically by a computer algorithm, or inferred through a relationship to another resource such as a hyperlink.
- *Metadata continues to accrue during the life of an information object or system.* Metadata is created, modified, and sometimes even disposed of at many points during the life of a resource.
- *One information object's metadata can simultaneously be another information object's data, depending on the kinds of aggregations of and dependencies between information objects and systems.* The distinctions between what constitutes data and what constitutes

metadata can often be very fluid and may depend on how one wishes to use a certain information object.

Why Is Metadata Important?

Metadata consists of complex constructs that can be expensive to create and maintain. How, then, can one justify the cost and effort involved? The development of the World Wide Web and other networked digital information systems has provided information professionals with many opportunities while at the same time requiring them to confront issues that they have not had occasion to explore previously. Judiciously crafted metadata, wherever possible conforming to national and international standards, has become one of the tools that information professionals are using to exploit some of these opportunities, as well as to address some emerging issues, discussed below.

Increased accessibility: Effectiveness of searching can be significantly enhanced through the existence of rich, consistent, carefully crafted descriptive metadata. Metadata can also make it possible to search across multiple collections or to create virtual collections from materials that are distributed across several repositories—but only if the descriptive metadata records are the same or can be mapped across all the collections. (Mary Woodley discusses this in more detail in the third chapter of this book.) Metadata standards that have been developed by different professional communities but include some common data elements (e.g. title, date, creator), such as CDWA Lite, Dublin Core, EAD, MARC XML, MODS, and TEI, are making it easier for users to negotiate between descriptive surrogates of information objects and digital versions of the objects themselves and to search at both the item and collection levels within and across information systems.[13]

Retention of context: Museum, archival, and library repositories do not simply hold objects. They maintain collections of objects that have complex interrelationships among themselves and a variety of associations with people, places, movements or styles, and events. In the digital world it is not unusual for a single object from a collection to be digitized and then for that digital surrogate to become separated from both its own cataloging information (descriptive metadata) and its relationship to the other objects in the same collection, resulting in a decontextualized information object. Metadata plays a crucial role in documenting and maintaining important relationships, as well as in indicating the authenticity, structural and procedural integrity, and degree of completeness of information objects. In an

[13] See, e.g., the LEADERS Project, http://www.ucl.ac.uk/leaders-project/index.htm.

archive, for example, by documenting the content, context, and structure of an archival record, metadata in the form of an archival finding aid is what helps to distinguish that record from decontextualized information.

Expanding use: Digital information systems for museum and archival collections make it easier to disseminate digital versions of unique objects to users around the globe who, for reasons of geography, economics, or other barriers, might otherwise not have an opportunity to view them. With new communities of users, however, come new challenges concerning how to make the materials most intellectually accessible. These new communities of users may have significantly different needs, language skills, and information-seeking behaviors from those of the traditional users for whom many existing information services were originally designed.

Learning metadata: Teachers, schoolchildren, and college students may want to search for and use information objects in quite different ways from those of scholarly researchers. Instructors may wish to develop lesson plans, or to scaffold learning so that students build on prior knowledge or are introduced to technical terminology. Specialized forms of metadata have been developed to address these needs.[14]

System development and enhancement: Metadata can document changing uses of systems and content, and that information can in turn feed back into systems development decisions. Well-structured metadata can also facilitate an almost infinite number of ways for users to search for information, to present results, and even to manipulate and to present information objects without compromising their integrity.

Multiversioning: The existence of information about, and surrogates of, cultural objects in digital form has heightened interest in the ability to create multiple and variant versions of information objects. This process may be as simple as creating both a high-resolution copy of a digital image for preservation or scholarly research purposes and a low-resolution thumbnail image that can be rapidly transferred over a network for quick reference purposes. Or it may involve creating variant or derivative forms to be used, for example, in publications, exhibitions, or schoolrooms. In either case, there must be metadata to relate the multiple versions of a given information object and to capture what is the same and what is different about each version. The metadata must also be able to distinguish what is qualitatively different in the various digitized versions or surrogates and the original physical object or item.

Legal issues: Metadata allows repositories to track the many layers of rights, licensing, and reproduction information that exist for original items as well as for their related information objects and the multiple

---

[14] See, e.g., Gateway to Educational Materials, http://www.thegateway.org/about/gemin-general/about-gem/; and IEEE 1484.12.1—2002 Standard for Learning Object Metadata.

versions of those information objects. Metadata also documents other legal or donor requirements that have been imposed on original objects and their surrogates—for example, privacy concerns, restrictions on reproductions, and proprietary and commercial interests. (See "Rights Metadata Made Simple," p. 63.)

Preservation and persistence: If digital information objects that are currently being created are to have a chance of surviving migrations through successive generations of computer hardware and software, or removal to entirely new delivery systems, they will need to have metadata that enables them to exist independently of the system that is currently being used to store and retrieve them. Technical, descriptive, and preservation metadata that documents how a digital information object was created and maintained, how it behaves, and how it relates to other information objects will be essential. It should be noted that for the information objects to remain accessible and intelligible over time, it will also be essential to preserve and migrate this metadata and to ensure that it does not become "disconnected" from the object that it describes.

System improvement and economics: Benchmark technical data, much of which can be collected automatically by a computer, is necessary to evaluate and refine systems in order to make them more effective and efficient from a technical and economic standpoint. The data can also be used in planning for new systems.

A Note on Metadata, Version Control, Reuse, and Recontextualization

It is worth giving special mention to the roles that metadata increasingly needs to play in supporting some of the particular opportunities of the digital age. Historically, one goal of cataloging was to make it possible to distinguish one version of an object or work from another. An item might be different, for example, because it was a second edition of the same work, because it contained distinctive printing anomalies from other copies printed at the same time, because it was an abridged or translated version of the original title, or because its title had changed.[15] Various standardized practices exist to help catalogers alert potential users to such differences in versions of a work. Today metadata must still be able to

[15] According to the FRBR conceptual model, these are different "expressions" and/or "manifestations" of a work. See http://www.ifla.org/VII/s13/frbr/frbr.htm. Note that the definition of a "work" (and the conceptual model) can differ considerably for unique works of art or architecture, as opposed to literary works or musical compositions, for which the FRBR model is ideal. See Murtha Baca and Sherman Clarke, "FRBR and Works of Art, Architecture, and Material Culture," in *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools*, ed. Arlene G. Taylor (Westport, CT: Libraries Unlimited, 2007), pp. 103–10.

elucidate such distinctions. However, it must also be able to help users distinguish between, and trace the changes in, the following:

- Original analog and digitized versions, noting any changes that might have occurred accidentally or deliberately during the digitization process (e.g., digital "repair" of a broken glass lantern slide).
- Digitized and born digital objects that are created in a range of resolutions to facilitate a variety of distribution mechanisms and uses, or that are periodically refreshed or migrated or rendered into an alternate format for preservation and long-term storage or security purposes.
- Original and renamed or retitled or reattributed objects. For example, museum objects may be renamed or reattributed or assigned a different creation date because new documentation has come to light. Metadata may also change due to cultural sensitivities or provenancial challenges; for example, place-names or object names may be changed to their original Native American forms, with English-language names assigned after the objects' creation "demoted" to the status of variants or additional access points.
- Original born digital materials and revised or updated versions (e.g., Web pages, reference databases).
- Original analog or born digital materials that are reused in part or in whole in new digital resources (e.g., personal Web pages, digital art, or digital music compilations).
- Objects, especially but not only museum objects, that are described collectively in one context within their metadata (e.g., as objects that were all collected at the same time at the same archaeological excavation) but are then taken individually out of that collection and recontextualized (e.g., in a special exhibition of Greek vases from a particular period or an exhibition of paintings relating to a particular theme or subject).

Conclusion and Outstanding Questions

Metadata is like interest: it accrues over time. To stretch the metaphor further, wise investments generate the best return on intellectual capital. Carefully crafted metadata results in the best information management—and the best end-user access—in both the short and the long term. If thorough, consistent metadata has been created, it is possible to conceive of it being used in an almost infinite number of new and even currently unforeseen ways to meet the needs of both traditional and nontraditional users,

for multiversioning, and for data mapping and mining. But the resources and intellectual and technical design issues involved in good metadata development and management are far from trivial. Some key questions that must be resolved by information professionals as they develop digital information systems and objects are:

- identifying which metadata schema or schemas should be applied in order to best meet the needs of the information creator, repository, and users. As mentioned above, selection of an inappropriate schema (e.g., EAD for museum collections that do not share a common provenance) serves neither the collection materials themselves nor the users who wish to find, understand, and use those materials. Also, in many cases, especially with complex objects or hierarchically structured archival and other types of collections, a combination of schemas working together (e.g., MARC and/or EAD at the collection level; MARC, Dublin Core, MODS, VRA Core, or CDWA Lite at the item level) may be the best solution.
- deciding which aspects of metadata are essential for the desired goal and how granular each type of metadata needs to be—in other words, how much is enough and how much is too much. There will likely always be important tradeoffs between the costs of developing and managing metadata to meet current needs and creating sufficient metadata that can be capitalized on for future, often unanticipated uses. Metadata creators should remember that good "core" metadata can be a valid approach both in economic and in intellectual terms; see Principles 2 and 7 of "Practical Principles for Metadata Creation and Maintenance," pp. 71-72.
- ensuring that the metadata schemas and controlled vocabularies, thesauri, and taxonomies (including folksonomies) being applied are the most up-to-date, complete versions of those sets of data values and that they are the appropriate terminologies for the materials being described and for the intended users.

What we do know is that the existence of many types of metadata will prove critical to the continued online and intellectual accessibility and utility of digital resources and the information objects that they contain, as well as the original objects and collections to which they relate. In this sense, metadata provides us with the Rosetta stone that will make it possible to decode information objects and their transformation into knowledge in the cultural heritage information systems of the future.