CHAPTER

METADATA

5

METADATA OVERVIEW

DEFINITIONS, STANDARDS, AND HISTORY

Metadata constitutes the foundation on which digital libraries are built. Commonly referred as to "data about data," metadata describes and organizes resources in the digital environment and enables users to discover and use the content of digital collections and repositories. Gilliland (2008) offers a broad definition of metadata as "the sum total of what one can say about any *information object* at any level of aggregation" (p. 2). Metadata can be used to capture characteristics and attributes of information resources on an item and/or collection level. The concept of metadata is used in diverse communities involved in organizing and managing information. In libraries and other cultural heritage institutions, the term is applied to the value-added information for arranging, describing, and enhancing intellectual access to information objects (Gilliland, 2008). Creation of high-quality metadata is essential to the access and preservation of digital library materials, including cultural heritage collections and scholarly publications in digital repositories.

Definitions of metadata in the library world emphasize the structured nature of metadata and the standardization of the metadata development process (NISO, 2004; Taylor and Joudrey, 2008; Zeng and Qin, 2008). NISO (2004) defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (p. 1). Zeng and Qin (2008) expand this definition by stating that metadata is "structured, encoded data that describes characteristics of information-bearing entities (including individual objects, collections, or systems) to aid in the identification, discovery, assessment, management, and preservation of the described entities" (pp. 321–322). The structure of metadata in digital libraries is governed by schemas. The concept of a metadata record grouping together all the statements about a resource is at the core of most schemas developed in digital libraries. Schemas are used in conjunction with other value and content standards, such as controlled vocabularies and input guidelines.

Libraries, archives, and museum communities have developed a range of different standards to guide the design, implementation, and exchange of structured metadata. Metadata is comprised of several building blocks including data structure standards and rules for formatting the contents of metadata records. Other standards determine how metadata is encoded and exchanged (Elings and Waibel, 2007; Miller, 2011; Mitchell, 2013a; Zeng and Qin, 2008). Thus, metadata standards are characterized by data structure (schemas), content, value, data format, and exchange:

- Schemas or metadata element sets are standards for data structures and semantics; the Dublin Core Metadata Element Set (DCMES) is an example of the most widely adopted schema in digital libraries.
- Data content standards provide the rules for metadata generation and formatting; *Anglo-American Cataloging Rules Revised,* Second Edition (AACR2), has been used in cataloging for many

Discover Digital Libraries. http://dx.doi.org/10.1016/B978-0-12-417112-1.00005-3

years and is currently being replaced by *RDA: Resource Description and Access*. Other content standards include *Describing Archives: A Content Standard* (DACS) and *Cataloging Cultural Objects* (CCO), which are used to describe archival materials and cultural objects.

- Data value standards are lists of standardized terms for recording values in metadata records. A number of authority files, subject headings, and thesauri are available from the Library of Congress and from the Getty Research Institute. The Library of Congress Subject Headings (LCSH) and the Art and Architecture Thesaurus (AAT) are examples of data value standards.
- Data format standards refer to standardized methods for encoding metadata so that computers can process data. Extensible Markup Language (XML) is a standard that is primarily used for encoding metadata in the digital library environment.
- Data exchange standards facilitate the sharing of metadata between collections and repositories. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is an interoperability standard for sharing metadata in the digital library environment. Open Archives Initiative Object Reuse and Exchange (OAI-ORE) is a standard for the description and exchange of aggregated of web resources, including compound digital objects (OAI, 2015). OAI-ORE is used in the aggregation of metadata in large-scale digital libraries and linked data projects (Mitchell, 2013c).

Elings and Waibel (2007) provide a review of the standards developed in libraries, archives, and museums and illustrate it with a grid of primary standards in use across different communities. The distinction between schemas and different standards for content and data exchange is also relevant in the linked data environment (Mitchell, 2013b).

Metadata practices in the digital library environment build on a strong tradition of cataloging and indexing in libraries, archives, and museums. In the print environment, books and other analog materials have been cataloged using MARC (MAchine-Readable Cataloging) as a structure standard and AACR2 and LCSH as data content and value standards. Many of these tools, like LCSH, have been adopted successfully for content description of digital objects. However, the application of MARC for describing complex and dynamic digital objects proved to be difficult due to its monolithic and inflexible structure. New schemas have been developed in the digital library environment to address the limitations of MARC and offer more flexibility in the means of describing and managing digital objects.

The term “metadata” can be traced back to the 1960s, but it became popular in database management literature in the 1980s (Lange and Winkler, 1997; Vellucci, 1998). Prior to the mid-1990s and the development of digital libraries, the term was used primarily by the communities engaged in the management of geospatial data and design of databases and systems (Gilliland, 2008). The early metadata initiatives in digital libraries include TEI and the Online Computer Library Center (OCLC) projects that resulted in the development of Dublin Core. The Text Encoding Initiative (TEI) published the Guidelines for Electronic Text Encoding and Interchange (TEI guidelines) in 1994; OCLC initiated the web resource cataloging project in 1994 by selecting AACR2 and the MARC format to catalog web materials. This led to the creation of the Dublin Core, which arose out of the Metadata Workshop held in 1995 (Zeng and Qin, 2008). Vellucci (1998) summarizes several metadata initiatives that surfaced in different communities in the 1990s. The library community applied traditional cataloging techniques to describe digital resources, such as MARC (Dillon and Jul, 1996). Simultaneously, scholarly, archival, and museum communities began using Standard Generalized Markup Language (SGML) or XML, such as in the TEI headers (Barbero and Trasselli, 2014; Beißwenger et al., 2012; Sperberg-McQueen, 1996; Sperberg-McQueen and Burnard, 1994).

The first decade of digital library development is marked by the explosion of metadata schemas, proposed to meet the needs of different communities and subject domains. The large number of standards, especially in the descriptive realm, demonstrates the need for individualized standards to correspond to varied contexts. The diversity of metadata standards reflects the evolving nature of information organization in digital libraries. Schema-based metadata represents two decades of intensive metadata development in the digital library environment. Linked data introduce a new data model and a significant shift in the way metadata is recorded and connected in the open web.

FUNCTIONS AND TYPES OF METADATA

Different types of metadata are needed for resource description, discovery, retrieval, use, presentation, and preservation of digital objects. The primary role of metadata is to identify, describe, and provide intellectual access to the content of a digital collection. As Miller (2011) notes, “without metadata, the collection would be virtually useless. Users would have no way to find and identify the digital objects within the collection” (p. 9). Metadata is particularly important in collections containing visual, sound, and moving image materials, which are very difficult to discover without textual description (Laursen et al., 2012).

In addition to facilitating resource discovery and use, metadata supports interoperability, organization, management, and preservation of digital objects (NISO, 2004). Lagoze et al. (1996) note the role of metadata in capturing terms and conditions of data, provenance, content rating data, linkage or relationship information, and structural data. Gilliland (2008) describes the role of metadata in maintaining the relationships between multiple versions of the same digital object and its role in retaining contextual information. The importance of recording data related to provenance, context of creation, and use is discussed in the context of digital preservation in Chapter 9.

Different types of metadata are recorded in digital libraries. The authors vary in the classification of metadata types, especially in regard to preservation and technical metadata that are often listed under administrative metadata. Gilliland-Swetland (1998) proposes a basic typology based on the functions of metadata:

- Administrative metadata that presents information associated with the management and organization of information resources
- Descriptive metadata that provides information to depict information resources
- Preservation metadata that offers information with respect to the conservation of information resources
- Technical metadata that illustrates information related to system functions and metadata behaviors

Miller (2011) classifies metadata into three categories: administrative, structural, and descriptive metadata, following a common approach adopted in library and information science literature:

- Descriptive metadata
  - elements describing or cataloging digital resources
  - information identifying the content of a digital item
  - terms required to retrieve a digital item or a group of digital items
- Administrative metadata
  - elements used for managing digital objects and collections

    - information life-cycle data from creation to dissemination
  - subtypes of administrative metadata
    - technical and preservation metadata
    - rights metadata
    - use metadata
- Structural metadata
  - elements offering a structure for a complex digital object or a group of associated digital objects
  - multiple files of one digital object (e.g., pages of a book)
  - multiple views of one digital object (e.g., different views of an object).

In this framework, technical and preservation metadata are listed as subtypes of administrative metadata, and structural metadata is a separate category.

Table 5.1 provides a summary of metadata types and their functions and lists a number of corresponding schema examples. As demonstrated in the table, the relationship between metadata types and schemas is not one-to-one. In fact, certain metadata schemas, such as Dublin Core or METS, can accommodate most of the types and include elements for recording descriptive as well as administrative metadata. In addition, there are also standards dedicated exclusively to technical and preservation metadata, such as MIX, based on the NISO technical standard for still images and new standards for capturing technical specifications of audio and video (AudioMD and VideoMD). More information about the technical standards is available at the Library of Congress Standards site (Library of Congress, 2011a, b). The digital preservation standard PREMIS is discussed in Chapter 9.

Table 5.1 Summary of Metadata Types and Their Functions

Metadata Type	Metadata Functions	Schema Examples
Descriptive	Describes an object; provides access points to facilitate resource discovery; indicates relationships	Dublin Core, MODS, EAD, VRA, CDWA
Administrative	Indicates ownership/digital provenance; provides management and rights information	METS, Dublin Core, VRA, EAD
Structural	Expresses the relationships of an object (or aggregation of objects) to other related objects; describes structural characteristics of compound objects	METS
Technical	Identifies digital objects and their technical specifications; certifies integrity and authenticity	MIX—NISO Metadata for Images, AudioMD, VideoMD
Preservation	Describes properties of digital objects in archival storage; records preservation activities	PREMIS, METS

METADATA SCHEMAS

Metadata schemas provide a foundation for structuring metadata. A schema is a predefined set of elements designed for a specific purpose, such as describing and managing information resources (NISO, 2004). The term "schema" is used to denote the singular form. The plural forms of schema are "schemas" or "schemata" (Baca, 2008). Schemas specify the names of elements and define their meaning. Some schemas may also provide the rules for how content must be formulated and encoded. Defined element sets are often represented in XML. Most of the schemas developed in digital libraries have a flat structure with a linear list of elements or a hierarchical structure indicating a parent–child relationship. A record remains a central concept of schemas developed in digital libraries where all attributes and characteristics of an information resource are grouped together. The underlying data models, however, are more flexible than MARC, allowing for the refinement of elements and the development of local application profiles.

Metadata schemas represent a significant departure from the traditional bibliographic description because of their varied and flexible approaches. Duval et al. (2002) outlines the fundamental principles that inform the design and application of metadata schemas:

- *Modularity* allows the combination of elements from different schemas, as well as vocabularies and other metadata building blocks, in a syntactically and semantically interoperable way.
- *Extensibility* allows the extension of a basic element set (repeat standard elements or add new ones) to accommodate local or domain-specific needs.
- *Refinement* allows the qualification of standard elements or an increased specificity of meaning; refinement also involves the specification of value sets or defining the range of values for a given element.
- *Multilingualism* addresses metadata design in light of linguistic and cultural diversity, with metadata schemas and records available to users in their native languages and in appropriate character sets.

Multiple metadata schemas have been developed in the cultural heritage communities to address the unique characteristics of resources in diverse knowledge domains. Digital objects are complex, often comprised of multiple files, and require extensive metadata for their management, use, and preservation. The wide array of metadata schemas also corresponds to the variety of formats, greater availability of audiovisual materials in the digital form, and finally, the different needs and traditions of cultural heritage organizations. Metadata schemas differ in:

- The underlying data model: flat structure (Dublin Core) or hierarchical (MODS or CDWA)
- The number of data elements
- Granularity of description
- The use of mandatory fields
- Encoding requirements
- The application of content rules and value standards

The following section provides an overview of the most frequently used schemas in digital libraries.

DUBLIN CORE

Dublin Core is one of the most widely adopted metadata schemas. It is used to meet the needs of a variety of user communities, including libraries, archives, museums, and other information providers

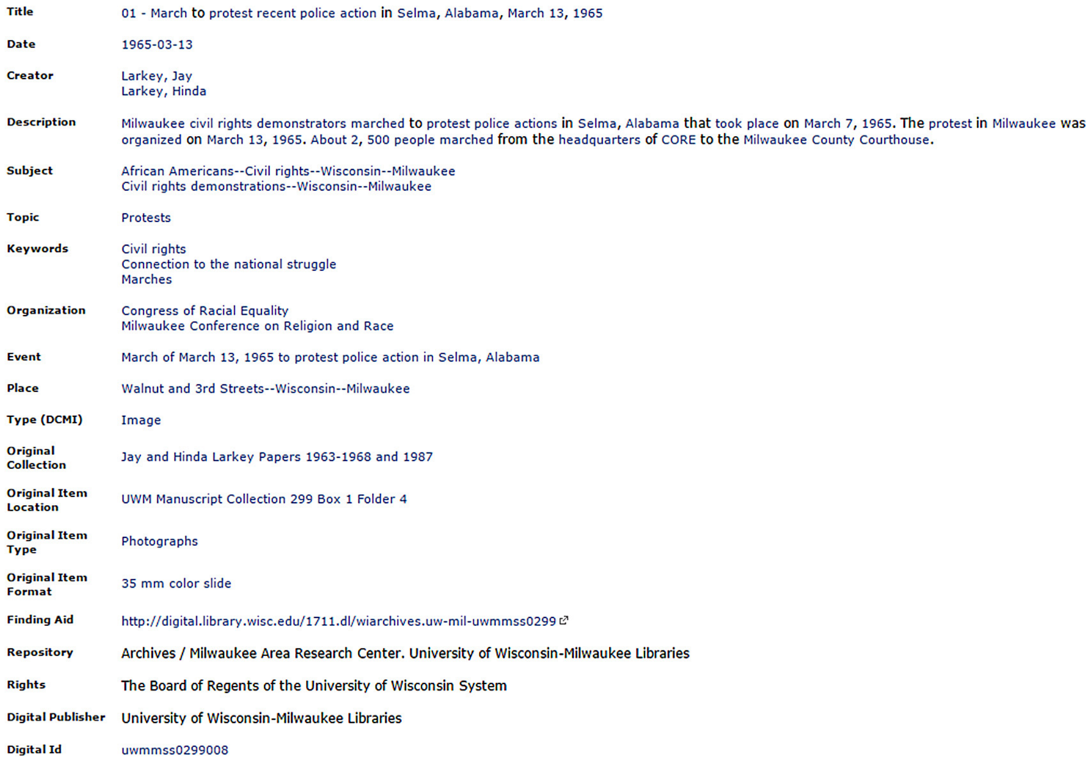
(Chan et al., 2001). The Dublin Core Element Set originated from a workshop that took place in 1995 in Dublin, Ohio, and hence the name. The Dublin Core Metadata Initiative (DCMI) was established a few years later. The first workshop was held to discuss how to deal with the need to describe and organize networked resources. A variety of working groups were created to advance the development of a new schema with the DCMI guiding the process. The mission of the DCMI is to facilitate the discovery of Internet resources by developing metadata standards, defining frameworks for the interoperability of metadata schemas, and assisting the creation of community- or discipline-specific metadata schemas (Weibel and Koch, 2000).

Dublin Core includes 15 original elements: contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, and type. In 2000, optional qualifiers were approved to enrich the schema. Beginning in 2012, the term "metadata qualifiers" was replaced by "refinements" and "encoding schemes" (DCMI, 2015a). Table 5.2 presents the basic Dublin Core Metadata Elements with associated definitions and examples. Fig. 5.1 shows an example of a customized Dublin Core record. The record example in Fig. 5.1 includes non-DC elements, such as keywords, event, and place, which have been mapped to Dublin Core. Fig. 5.1 presents a public display of a DC record. Metadata mapping is conducted by library professionals in the administrative module of the underlying software.

Table 5.2 Dublin Core (Version 1.1)

Elements	Definition
Contributor	An entity responsible for making contributions to the resource
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant
Creator	An entity primarily responsible for making the resource
Date	A point or period of time associated with an event in the lifecycle of the resource
Description	An account of the resource
Format	The file format, physical medium, or dimensions of the resource
Identifier	An unambiguous reference to the resource within a given context
Language	The language of the resource
Publisher	An entity responsible for making the resource available
Relation	A related resource
Rights	Information about rights held in and over the resource
Source	A related resource from which the described resource is derived
Subject	The topic of the resource
Title	A name given to the resource
Type	The nature or genre of the resource

Dublin Core Metadata Element Set (http://dublincore.org/documents/dces/)



Title	01 - March to protest recent police action in Selma, Alabama, March 13, 1965
Date	1965-03-13
Creator	Larkey, Jay Larkey, Hinda
Description	Milwaukee civil rights demonstrators marched to protest police actions in Selma, Alabama that took place on March 7, 1965. The protest in Milwaukee was organized on March 13, 1965. About 2, 500 people marched from the headquarters of CORE to the Milwaukee County Courthouse.
Subject	African Americans--Civil rights--Wisconsin--Milwaukee Civil rights demonstrations--Wisconsin--Milwaukee
Topic	Protests
Keywords	Civil rights Connection to the national struggle Marches
Organization	Congress of Racial Equality Milwaukee Conference on Religion and Race
Event	March of March 13, 1965 to protest police action in Selma, Alabama
Place	Walnut and 3rd Streets--Wisconsin--Milwaukee
Type (DCMI)	Image
Original Collection	Jay and Hinda Larkey Papers 1963-1968 and 1987
Original Item Location	UWM Manuscript Collection 299 Box 1 Folder 4
Original Item Type	Photographs
Original Item Format	35 mm color slide
Finding Aid	http://digital.library.wisc.edu/1711.dl/wiarchives.uw-mil-uwmmss0299
Repository	Archives / Milwaukee Area Research Center. University of Wisconsin-Milwaukee Libraries
Rights	The Board of Regents of the University of Wisconsin System
Digital Publisher	University of Wisconsin-Milwaukee Libraries
Digital Id	uwmmss0299008

FIGURE 5.1 An Example of a Customized Dublin Core Record

http://collections.lib.uwm.edu/cdm/singleitem/collection/march/id/1531/rec/1

Dublin Core can be further enhanced by element refinements and element encoding schemes (DCMI, 2015a, b; Miller, 2011). Examples of refinements to the Date element are created, valid, available, issued, and modified. Another example can be seen in refinements to the Relation element, which include Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, and Has Format. Element encoding schemes offer either a controlled vocabulary scheme or a standard syntax encoding scheme to associate an element with an existing controlled vocabulary, formal notation, or set of rules. This helps increase the precision of information retrieval. Examples of element encoding schemes for subjects are LCSH, MeSH, TGM, or AAT. Examples of element encoding schemes for language are ISO 638-2, ISO 639-3, RFC 1766, and RFC 4646. Flexibility is one of the main characteristics of Dublin Core. There are no required elements and each element can be used multiple times. Simplicity and semantic interoperability are the two key attributes (Chan et al., 2001).

The Dublin Core schema has been widely adopted in the digital library environment. Dublin Core has been implemented not only domestically in the United States but also internationally. The two main obstacles to its adoption are the relative scarcity of elements and qualifiers and an insufficient number of guidelines for its use. Even though Dublin Core is considered easy to use, Chuttur (2014) discovered a high error rate across the groups that used best practice guidelines and definitions, though using the guidelines resulted in fewer errors. The development and adoption of updated best practice guidelines based on user needs is essential to generating useful metadata records.

Implementing Dublin Core in repositories, where content contributors are responsible for providing metadata elements, brings a new set of challenges. In one study of using this schema in DSpace, an open source institutional repository software, the results reveal different quality issues, such as incomplete records caused by skipping nonmandatory fields, a lack of authority control over subject headings, low metadata accuracy caused by unclear element definition, and metadata inconsistency due to a lack of required conventions (Kurtz, 2013).

METADATA OBJECT DESCRIPTION SCHEMA (MODS)

Metadata Object Description Schema (MODS) is a schema that has its roots in bibliographic description. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress (Library of Congress, 2015a). MODS is derived from MARC 21. As such, it is highly compatible with MARC fields. The first version of MODS was developed in 2002 and based on feedback on the first version; version 2.0 was published in 2003. Version 3.0 was made available in late 2003.

The structure of MODS is hierarchical. It contains parent elements and child elements. Attributes are used simultaneously to refine the meaning of an element. McCallum (2004) specifies several requirements and characteristics of MODS: developed for an XML environment, compatible with MARC21, simple, and having a modest amount of top-level elements. The following features correspond to the requirements and characteristics of MODS: user-oriented tags, regrouped data elements in MARC, fewer coded values, added electronic resource data, linking flexibility, recursion for related items, special attributes, round-trip transformation with MARC21, and allowance for mixed content. In addition, one characteristic of MODS is that it offers users the opportunity to use different levels of granularity. (Miller, 2011). MODS includes 20 top-level elements: titleInfo, name, typeOfResource, genre, originInfo, language, physicalDescription, abstract, tableOfContents, targetAudience, note, subject, classification, relatedItem, identifier, location, accessCondition, part, extension, and recordInfo (Library of Congress, 2014a). In addition, MODS has 47 subelements (Guenther, 2003). Table B.1 in Appendix B presents the MODS top-level elements with associated definitions.

The benefits of using MODS include its hierarchical structure allowing for granular description, detailed user guidelines, and mappings with examples. More importantly, transformation tools assisting in conversion from MODS to MARC and other metadata schemas are also available (Guenther, 2004). As Dulock (2012) describes, MODS is used for descriptive metadata in digitization projects because of its ample metadata element set. However, MODS metadata needs to be mapped to Dublin Core for harvesting purposes because of the requirements of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Dulock, 2012).

METADATA ENCODING AND TRANSMISSION STANDARD (METS)

METS is sponsored by the Digital Library Federation, supported by the Library of Congress, and governed by the METS Editorial Board. Its first schema was made available in 2001. The development of

METS continues today (Library of Congress, 2015b). METS is defined as "an XML schema designed for the purpose of creating XML document instances that express the hierarchical structure of digital library objects, the names and locations of the files that comprise those digital objects, and the associated descriptive and administrative metadata" (Cundiff, 2004, p. 53). The main user communities of METS are university libraries, archives, and museums. In general, a METS document can have seven major subsections:

- Mets Header (metsHdr)
- Descriptive Metadata Section (dmdSec)
- Administrative Metadata Section (amdSec)
- File Section (fileSec)
- Structure Map (structMap)
- Structural Links (structLink)
- Behavior Section (behaviorSec)

Three types of metadata are identified in METS: descriptive metadata, administrative metadata, and structural metadata. METS does not offer its own vocabularies for descriptive metadata. Instead, it offers three options: (1) point to metadata in external documents; (2) point to systems using the Metadata Reference (mdRef) element; and (3) embed descriptive metadata using the Metadata Wrap (mdWrap) element. Similarly, METS does not offer its own vocabularies for administrative metadata. The administrative metadata section can be further classified into four subsections: technical metadata (techMD), rights metadata (rightsMD), source metadata (sourceMD), and digital provenance metadata (digiprovMD) (Cantara, 2005; Cundiff, 2004). The structural map is the core of a METS document, and it is the only required subsection. It represents the hierarchical structure and sequence of components of a digital object in the form of nested divisions of elements. It comprises attributes such as ID, LABEL, TYPE, and ORDER. Table B.2 in Appendix B presents the METS elements with associated definitions.

METS has been implemented in a variety of digital library projects. The METS Implementation Registry is available at the Library of Congress Standards site (Library of Congress, 2015b). METS was selected for the University of Texas' Human Rights Documentation Initiative (HRDI) because of the schema's capability in managing digital objects and metadata at several levels (Dulock, 2012). METS functions as a wrapper in which the descriptive metadata and administrative metadata sections can be connected to other parts of the METS document.

METS facilitates interoperability, but its flexibility in representing digital library objects also poses some challenges. McDonough (2006) highlights some issues with METS in terms of interoperability:

- Allows users to insert arbitrary metadata schema
- Does not control the location of metadata or the associated format
- Offers no guarantee in applying standard rules of description
- Lacks controlled vocabularies for the attributes of some elements
- Lacks structural constraints on a digital object

METS profiles enable institutions to record restrictions and requirements for the compiling of METS documents. Although METS profiles do not guarantee the interoperability, the profiles are the starting point from which to consider issues regarding sharing and exchanging complex digital objects. It is worth noting that other metadata schemas for complex digital objects share similar problems.

TEXT ENCODING INITIATIVE (TEI)

TEI, a standard for encoding textual documents, was developed as one of the first digital library metadata standards. A conference at Vassar College in Poughkeepsie, New York, in 1987 proposed the Poughkeepsie Principles, which set up the foundation of TEI. In 1990, the first edition of TEI was published, and it adopted SGML for the coding. SGML is an international standard for document markup. The most notable edition is the P3 *Guidelines for Electronic Text Encoding and Interchanges*, which was made available in 1994. It defines 600 elements for the ending of text. The P4 version was publicized in 2001 by the TEI Consortium, when TEI started to support both XML and SGML. The P5 edition is an XML-based markup for digital texts. TEI has concentrated more on functional aspects of texts (Ore and Eide, 2009; Vanhoutte, 2004). Several working groups contributed to the production of TEI P5. According to Cummings (2008), "The TEI Guidelines are not only a guide to best practice, but are also an evolving historical record of the concerns of the field of Humanities Computing" (p. 1). In that case, the TEI guidelines not only show the evolution of the recommendations but also the influences of the technical and theoretical background and development.

TEI P5 consists of 23 chapters organized from broad to specific topics. The first part (5 chapters) introduces TEI to potential users. The second part (7 chapters) focuses on each kind of specific text: verse, drama, spoken text, dictionaries, and manuscript materials. The third part (9 chapters) takes care of topics that are associated with specific applications. The fourth part (2 chapters) discusses how to encode the XML to represent the TEI scheme. The last part (1 chapter) concentrates on TEI customization and conformance. The following are the three primary functions of the TEI guidelines:

- Guide individual or local practice in text creation and data capture
- Support data interchange
- Support application-independent local processing (TEI Consortium, 2015)

The TEI encoding schema consists of a number of modules including specific XML elements and their attributes. In principle, a TEI schema may contain any combination of different modules. Among all modules, four modules are of particular importance:

- The TEI module specifies classes, macros, and data types.
- The core module comprises declarations for elements and attributes.
- The TEI header module presents declarations for the metadata elements and attributes.
- The text structure module is required for the encoding of most book-like objects (TEI Consortium, 2015).

These four modules are used in almost all the TEI schemas. TEI also defines several hundred elements and attributes. Each definition contains the following components (TEI Consortium, 2015, p. 1):

- A prose description
- A formal declaration, expressed by special-purpose XML vocabulary combined with elements extracted from the ISO schema language RELAX NG
- Usage examples

As an example, Table B.3 in Appendix B presents the TEI Header elements with associated definitions and examples.

TEI was selected to promote the exchange of data among national and international projects by exporting manuscript descriptions from a library system into TEI XML documents (Barbero and

Trasselli, 2014). TEI allows customization of elements to satisfy the needs of a variety of digital projects. Although customization is one of its design goals, that feature also creates potential problems with data sharing and exchange (Cummings, 2008). Customization is an effort to apply the TEI encoding framework to new genres and document types. Beißwenger et al. (2012) demonstrate that TEI Guidelines can be customized and applied to different types of Computer Mediated Communication (CMC) genres. TEI has mostly been applied in large projects in the humanities and social sciences. It is challenging for small institutions to implement TEI because of the detailed requirements for text encoding. The benefits of enhanced access to the digital collection may outweigh the problems in certain applications (Wisneski and Dressler, 2009).

ENCODED ARCHIVAL DESCRIPTION (EAD)

Encoded Archival Description (EAD) is an international standard for encoding finding aids for archival materials, with version 1.0 published in 1998 and revised in 2002. The standard originated from a research project at the University of California at Berkeley. Just like TEI, it was originally an application of SGML, and it became XML compatible later. EAD is maintained by the Network Development and MARC Standards Office of the Library of Congress in collaboration with the Society of American Archivists (Library of Congress, 2013). McCrory and Russell (2013) state, "EAD provides a means of structuring the language of finding aids so that they may be processed for presentation on the web and so that their descriptive elements can be exchanged with other metadata systems" (p. 99).

The EAD record structure consists of three groups of categories: <eadheader> (EAD Header), <frontmatter> (Front Matter), and <archdesc> (Archival Description). Among them, <eadheader> and <archdesc> are required elements. Multilevel description is one of the key features of EAD. The first level describes the collection; the second level presents a series of materials; the third level illustrates the individual folder. An EAD record contains elements, associated attributes, and allowed values. EAD encoders need to record the archival collections as a whole, and more importantly, they also need to provide the descriptive data for each series, box, folder, and item (Zeng and Qin, 2008). Two processes are involved in the adoption of EAD: encoding to match information in a finding aid to EAD elements and publication to make a finding aid available on the web (Yakel and Kim, 2005). EAD conforms to both SGML and XML specifications. The implementation of EAD is a complicated process because of the many options and models that need to be considered. Even though it does not necessarily reflect the exact structure of the finding aids produced, it does provide elements to represent the captured information. Table B.4 in Appendix B presents the EAD header elements with associated definitions and examples.

EAD was developed specifically to describe archival materials. Accordingly, the EAD structure adheres to the structure of an archival collection. In EAD, each level of the structure and its corresponding metadata are associated with each other (Niu, 2013). Although EAD is widely implemented in the archival field today, in a study on the diffusion and adoption of EAD in the archival community, Yakel and Kim (2005) found that the adoption rate was low, with only 42% of the respondents implementing EAD in their programs. Based on a usability study of an EAD interface, Yakel (2004) identifies several barriers for implementing EAD. The main barriers include the following: (1) Users are not familiar with EAD jargon, and (2) users do not understand the hierarchical structure of EAD. Yaco's (2008) survey findings yield similar results. After conducting a usability study on an EAD finding aid, DeRidder et al., (2012) recommend the following to enhance the usability of EAD finding aids: adding a navigational frame, adding a "search in page" feature, and modifying archival terminology.

VISUAL RESOURCES ASSOCIATION (VRA) CORE

The Visual Resources Association (VRA) Core is a metadata schema for visual resources, used primarily in art libraries and museums. According to Eustis (2013), "VRA's primary aim was to allow its members to collaborate on best practices for creating, describing, and distributing digital objects for resources such as images or cultural artifacts" (p. 441). It has also been implemented into digital library software, such as CONTENTdm. The first set of VRA Core elements was first released in 1996. The VRA 4.0 version was released in 2007. The standard is hosted by the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the VRA (Library of Congress, 2014b). The VRA Core consists of the following 19 elements: record type (collection/work/image), agent, cultural context, date, description, inscription, location, material, measurements, relation, rights, source, stateEdition, stylePeriod, subject, technique, textref, title, and worktype. Additionally, nine global attributes are used to define each element or subelement. They include dataDate, extent, href, pref (preferred value), refid (link to internal identifiers), rules, source, vocab, and xml:lang (Visual Resources Association, 2007). Table B.5 in Appendix B presents the VRA Core 4.0 elements with associated definitions and examples.

According to a 2011 survey, 56 institutions implemented the VRA Core schema (Mixter, 2014). Van Assem et al. (2010) discuss their decision in selecting VRA Core 3.0 for the MultimediaN E-Culture project. The main reasons are that (1) VRA Core elements map well to the raw data; (2) there is clear link between the VRA Core and the Dublin Core; and (3) the VRA Core offers a coherent set of initial facets for a facet browser. There are, however, some issues in mapping richer VRA Core elements to basic Dublin Core. Even though the data elements are similar, there are semantic differences between these elements. For example, the VRA Style/Period is similar to the Dublin Core elements subject or coverage, but some loss of meaning occurs if Style/Period is reduced to either one of the Dublin Core elements (Attig et al., 2004).

CATEGORIES FOR THE DESCRIPTION OF WORKS OF ART (CDWA)

Categories for the Description of Works of Art (CDWA) is a schema that originated in the art museum community. It is designed specifically for developing a structured approach to describing works of art, architecture, and other material culture. CDWA was developed in the late 1990s by the Art Information Task Force (AITF), a group of representatives from the art library and museum communities. The work of the Task Force was partially funded by the J. Paul Getty Trust College Art Association, and the standard as well as the implementation guidelines are available at the Getty's web site (Baca and Harpring, 2009).

CDWA provides a broad framework from which existing art information can be mapped and upon which new systems can be developed or linked. It also identifies vocabulary tools and provides guidelines for their use (Baca and Harpring, 2009). CDWA allows a greater level of granularity than Dublin Core, MODS, or even VRA. For example, it defines a number of categories for Creator, such as Creator Description, Creator Identity, and Creator Role. It also includes a record for the relationship between the object and its visual and textual representation. The full set of metadata elements is quite extensive and includes 540 categories and subcategories. Within the set, a number of categories are identified as "core"—considered necessary to describe a work of art. The schema is built on a hierarchical parent–child data model with subcategories nested under the main categories. The CDWA schema is used in combination with a content standard, CCO.

CDWA Lite is a subset of the full CDWA element set. It is based on the core elements of CDWA and the guidelines included in CCO. CDWA Lite is encoded in XML. It includes 22 core elements, with several nested subelements. The purpose of CDWA Lite is to provide a structured format for sharing core records of works of art and cultural materials between museums and other repositories of visual art. CDWA Lite is compatible with the Open Archives Initiative (OAI) harvesting protocol. Woodley (2008) describes a case study of the sharing of CDWA Lite-based metadata records of images of European tapestries from the Getty Research Institute Photo Study Collection and harvesting them into ARTstor.

Several metadata schemas are described and discussed earlier. Multiple schemas have been developed for different purposes and audiences. There is no single metadata schema that could fully represent descriptive, preservation, and structural metadata. In practice, several schemas have to be used to capture different types of metadata.

INTEROPERABILITY: METADATA MAPPING AND HARVESTING

Interoperability refers to the ability of multiple systems with different hardware and software, data structures, and interfaces to exchange and share metadata (NISO, 2004). The digital library environment, with multiple schemas and content and value standards, poses many challenges for metadata exchange across collections and repositories. The goal of interoperability is to enable the exchange of data between digital library systems and to provide services that simplify discovery and increase interactions with digital library resources in a network environment (Arms et al., 2002). Metadata mapping tools and shared transfer protocols have been developed to advance interoperability and improve access to digital library resources through metadata harvesting and large-scale repositories of aggregated metadata.

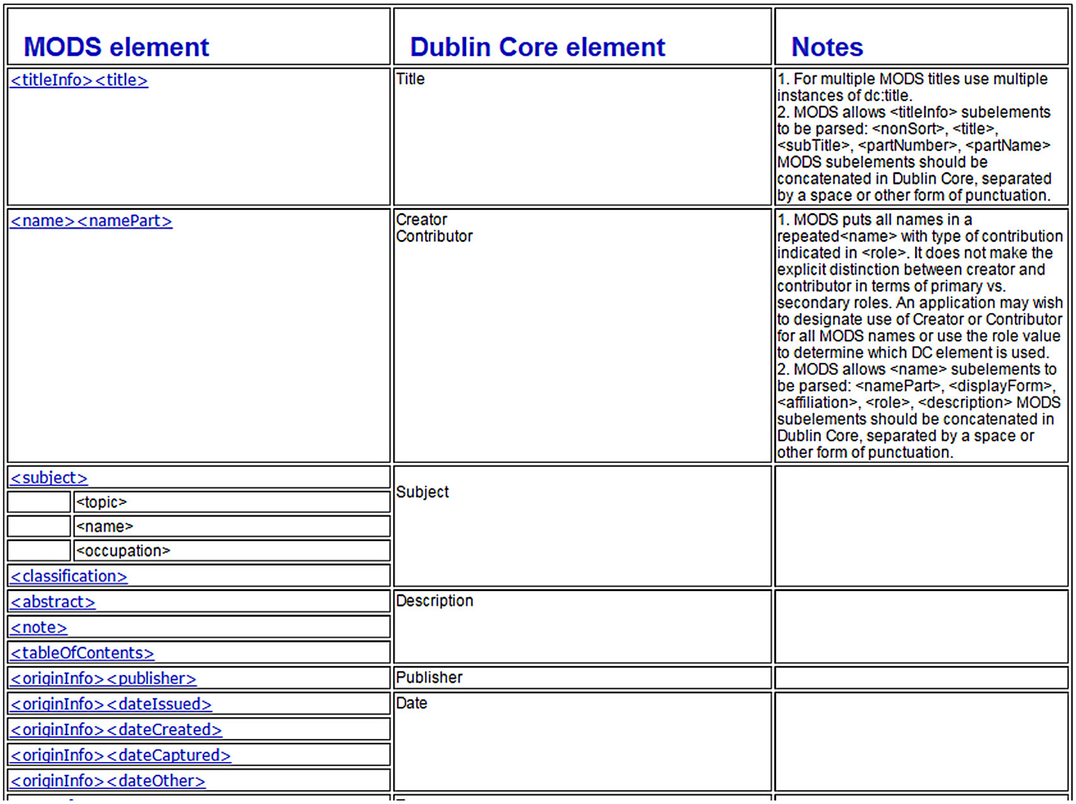
Metadata mapping facilitates the exchange of metadata between collections and systems using different schemas. The terms "metadata mapping" and "metadata crosswalk" are often used interchangeably (Woodley, 2008). Woodley (2008) offers a distinction between the terms by defining mapping as "the intellectual activity of comparing and analyzing two or more metadata schemas," while crosswalks are the products of the mapping activity and can be represented as tables or charts (p. 40). Interoperability between schemas can be examined on several levels, including semantic, structural, and syntactic. Most mapping activities in digital libraries focus on semantics analyzing the definitions of the elements in two or more schemas to determine whether they have the same or similar meanings and deciding which element in one schema can be mapped to an equivalent element in the second schema. Typically, mapping is performed in preparation for the exchange of metadata between systems but can also be done during metadata design, when elements of a local application profile need to be mapped to a standardized schema. Metadata mapping ensures cross-collection searching and exposing the local metadata to a wider audience. Mapping of customized schemas to a standard element set is further discussed in the section "Designing and Implementing Metadata."

Crosswalks provide specifications for mapping one metadata schema to another and assist in converting metadata created by different communities to be included in digital libraries and shared repositories. Several crosswalks have been developed to facilitate the mapping of popular schemas. The LC provides access to a number of crosswalks for the schemas that are maintained by LC, including MARC, MODS, and EAD. These schemas are mapped to Dublin Core but also to other standards. For example, MODS is mapped to Dublin Core and MARC, and the conversion is bidirectional—a crosswalk is also available for MARC to MODS and Dublin Core to MODS (Library of Congress, 2015c).

The Getty Research Institute (2014) published a crosswalk chart for multiple schemas, including CDWA, CDWA Lite, and VRA Core.

Semantic mapping is rarely direct, due to a different number of elements and different levels of granularity among schemas. The mapping of a richer schema to a simpler set of elements usually results in some loss of information. Miller (2011) points out that "mapping from one schema to another is virtually impossible without metadata degradation" (p. 233). Fig. 5.2 provides a sample of mapped elements from MODS to Dublin Core. This example demonstrates that some refinements of elements, such as subject, will be lost when MODS is mapped to Dublin Core.

Metadata harvesting has been developed as an approach to interoperability and metadata sharing between digital collections and repositories. This method addresses the difficulties of resource discovery in the digital library environment by gathering metadata from individual digital collections and providing access through an aggregated platform. The transfer of metadata is defined by the OAI-PMH (OAI, 2015). This metadata exchange standard requires data providers to expose their metadata as a set of simple Dublin Core. If original metadata is built with other schemas, such as MODS, Qualified



MODS element		Dublin Core element	Notes
<titleInfo><title>		Title	1. For multiple MODS titles use multiple instances of dc:title. 2. MODS allows <titleInfo> subelements to be parsed: <nonSort>, <title>, <subTitle>, <partNumber>, <partName> MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation.
<name><namePart>		Creator Contributor	1. MODS puts all names in a repeated<name> with type of contribution indicated in <role>. It does not make the explicit distinction between creator and contributor in terms of primary vs. secondary roles. An application may wish to designate use of Creator or Contributor for all MODS names or use the role value to determine which DC element is used. 2. MODS allows <name> subelements to be parsed: <namePart>, <displayForm>, <affiliation>, <role>, <description> MODS subelements should be concatenated in Dublin Core, separated by a space or other form of punctuation.
<subject>		Subject	
	<topic>		
	<name>		
	<occupation>		
<classification>			
<abstract>		Description	
<note>			
<tableOfContents>			
<originInfo><publisher>		Publisher	
<originInfo><dateIssued>		Date	
<originInfo><dateCreated>			
<originInfo><dateCaptured>			
<originInfo><dateOther>			

FIGURE 5.2 Mapping of Selected MODS Elements to Dublin Core (Library of Congress, 2015b)

A full set of mapped elements is available at: http://www.loc.gov/standards/mods/mods-dcsimple.html.

Dublin Core, TEI, or VRA, mapping and transformation of metadata to simple Dublin Core need to take place prior to harvesting. The OAI-PMH standard is discussed in more detail in Chapter 6.

Interoperability is a vital issue in digital library research and practice. In Lopatin (2010), OAI-PMH emerges as the most widely adopted solution to interoperability for both academic and nonacademic libraries, although more academic libraries (77%) than nonacademic libraries (69%) selected it. The study by Park and Tosaka (2010) yields similar results. About 36.8% of the respondents exposed their metadata through OAI harvesters. Metadata interoperability is a challenge for many of the institutions because of the financial barriers, personnel requirements, and technical constraints.

DESIGNING AND IMPLEMENTING METADATA

The process of designing and implementing metadata in digital libraries is highly structured, but it also involves a significant amount of customization. The development of local, collection-specific metadata is achieved through a wide range of schemas and their modular character. As Duval et al. (2002) note, "in a modular metadata world, data elements from different schemas as well as vocabularies and other building blocks can be combined in a syntactically and semantically interoperable way" (p. 2). This flexible approach is quite different from the uniformity of bibliographic description that is found in traditional library catalogs. The use of multiple customizable metadata schemas allows one to address different user needs, unique characteristics of collections, and diverse disciplinary domains. However, the lack of a uniform metadata model, and varied metadata creation practices poses challenges to quality, interoperability, and metadata sharing and reuse (Hillmann, 2008; Park, 2009; Park and Tosaka, 2010).

The multiple roles that metadata plays in digital libraries complicate the process of designing consistent and interoperable metadata. As discussed at the beginning of this chapter, different kinds of metadata are needed for resource description, discovery, use, presentation, and preservation of digital objects. Different types are related to multiple functions of metadata that go beyond the description of digital objects. The presence of multiple goals and corresponding data types may require more than one metadata set to be associated with an object, which further complicates the process of designing, implementing, and maintaining metadata in practical digital libraries.

The role of metadata in creating and preserving digital objects is underscored in *The Framework of Guidance for Building Good Digital Collections,* an NISO guide to recommended digital library practice (NISO Framework Working Group, 2007). *The Framework* defines basic requirements for metadata and outlines six general principles for designing and implementing high-quality metadata:

1. Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.
2. Good metadata supports interoperability.
3. Good metadata uses authority control and content standards to describe objects and collocate related objects.
4. Good metadata includes a clear statement of the conditions and terms of use for the digital object.
5. Good metadata supports the long-term curation and preservation of objects in collections.
6. Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification (NISO Framework Working Group, 2007, pp. 61–62).

These principles build on the traditions of library cataloging, especially regarding the adherence to standards and use of authority control and content standards. They also highlight the new roles of metadata in supporting interoperability, rights management, and long-term preservation. The emphasis is on the standardization of the metadata creation process, which in turn supports consistent and accurate resource description, interoperability, and the preservation of digital objects. As indicated in principle no. 6, metadata records themselves are digital objects and should have the attributes of good objects in order to be maintained and preserved.

The digital library environment offers a wide range of structure and content standards to support consistent metadata creation and interoperability. The metadata design process requires a number of decisions about the selection and integration of different standards and tools. As discussed in the previous section, the digital library environment offers multiple schemas, developed by different communities that are intended to address the unique characteristics of materials in diverse knowledge domains. Schemas are used in combination with other metadata building blocks, such as authority control tools and content standards. Library, archival, and museum communities provide a range of general and discipline-specific controlled vocabulary tools to record authorized forms of names and consistent subject terms. Communities also utilize content input guidelines to ensure consistent data formatting with regard to syntax, punctuation, capitalization, etc. Many of the content tools were originally developed in the print environment and have been adopted for metadata creation in digital libraries.

Table 5.3 provides a summary of the standards used for metadata creation in digital libraries. The list was compiled based on the results of studies that surveyed metadata practices in digital collections and repositories (Lopatin, 2010; Moulaison et al., 2015; Palmer et al., 2007; Park and Tosaka, 2010). There are some discrepancies between the studies when identifying the most frequently used schemas. Palmer et al. (2007) conducted a longitudinal study of IMLS digital collections created between 2003 and 2006 and found that Dublin Core and MARC were the two top metadata schemas used during that period. Lopatin (2010) and Moulaison et al. (2015) find that Dublin Core is the most widely adopted standard in digital collections and repositories, while the results of Park and Tosaka's (2010) survey indicate that MARC is still the most frequently used standard, followed by Dublin Core. Because of the disparate results, the standards in Table 5.3 are listed alphabetically, rather than in the ranking order.

The schemas and data value standards were selected for inclusion in Table 5.3 if they were listed in at least two of the three reviewed studies. Data content standards are discussed only in the Park and Tosaka (2010) study. The list of standards and tools is not exhaustive. The schemas identified in the reviewed studies focus on descriptive metadata. Interestingly, none of the studies mentions the structural metadata standard METS, nor do they mention the preservation standard PREMIS. Lopatin (2010) and Park and Tosaka (2010) mention a significant use of home-grown schemas and locally developed

Table 5.3 Metadata Standards Used in Digital Collections and Repositories

Type of Standard	Standard Name
Metadata schemas	Dublin Core; CDWA; EAD; MARC; MODS; Qualified Dublin Core; TEI; VRA
Data value standards/controlled vocabularies	AAT; LCSH; LC NAF; LC TGM; MESH; TGN
Data content standards	AACR2; CCO; DACS; DCRM; Dublin Core guidelines; EAD guidelines; LC MODS guidelines

vocabularies and guidelines. Park and Tosaka (2010) record that 25.1% of the survey participants engaged in creating local metadata elements and home-grown content guidelines. The authors discuss the problematic nature of this approach insofar as it hinders interoperability and metadata sharing in distributed environments.

The research studies confirm that multiple schemas and content standards are indeed used in digital libraries. This multiplicity is rooted in the different traditions of describing and organizing resources in cultural heritage communities. The schemas that originated in archives and museums tend to be used with other standards and tools developed in archival and visual resources communities. For example, EAD, a finding aid standard for archival collections, is often used in conjunction with DACS, which is a content standard of archival practice. VRA Core and CDWA are two structure standards designed specifically for creating metadata for works of art. CDWA is used in conjunction with CCO, a museum data standard for describing works of art and material culture, and with controlled vocabulary tools developed by the Getty Research Institute, including the *Art and Architecture Thesaurus* (AAT) or Thesaurus of Geographic Names (TGN). The VRA schema is also used with CCO and Getty vocabularies (Elings and Waibel, 2007). Dublin Core and MODS are both cross-disciplinary and general digital library standards, and are used with a variety of controlled vocabularies, including LC Subject Headings, Thesaurus for Graphic Materials (TGM), and Getty vocabularies.

The creation of metadata in digital collections and repositories involves two distinct phases:

1. Conceptual work in metadata design
2. Resource-intensive construction of metadata records

Metadata design, which includes selecting and customizing a schema, provides a foundation for building metadata records. Schema selection and/or the development of a local application profile takes place in the beginning phase of a digital library project. Metadata design is critical to the subsequent stages of metadata implementation and interoperability. This phase involves not only selecting an appropriate metadata schema but also determining the appropriate level of description and identifying controlled vocabulary tools to be used in creating records. The decision of whether to adopt an established content standard or develop local input guidelines is also made in the planning phase. Metadata implementation involves building records for digital objects based on the standards established in the design phase, executing quality control, and producing documentation.

SELECTING A SCHEMA

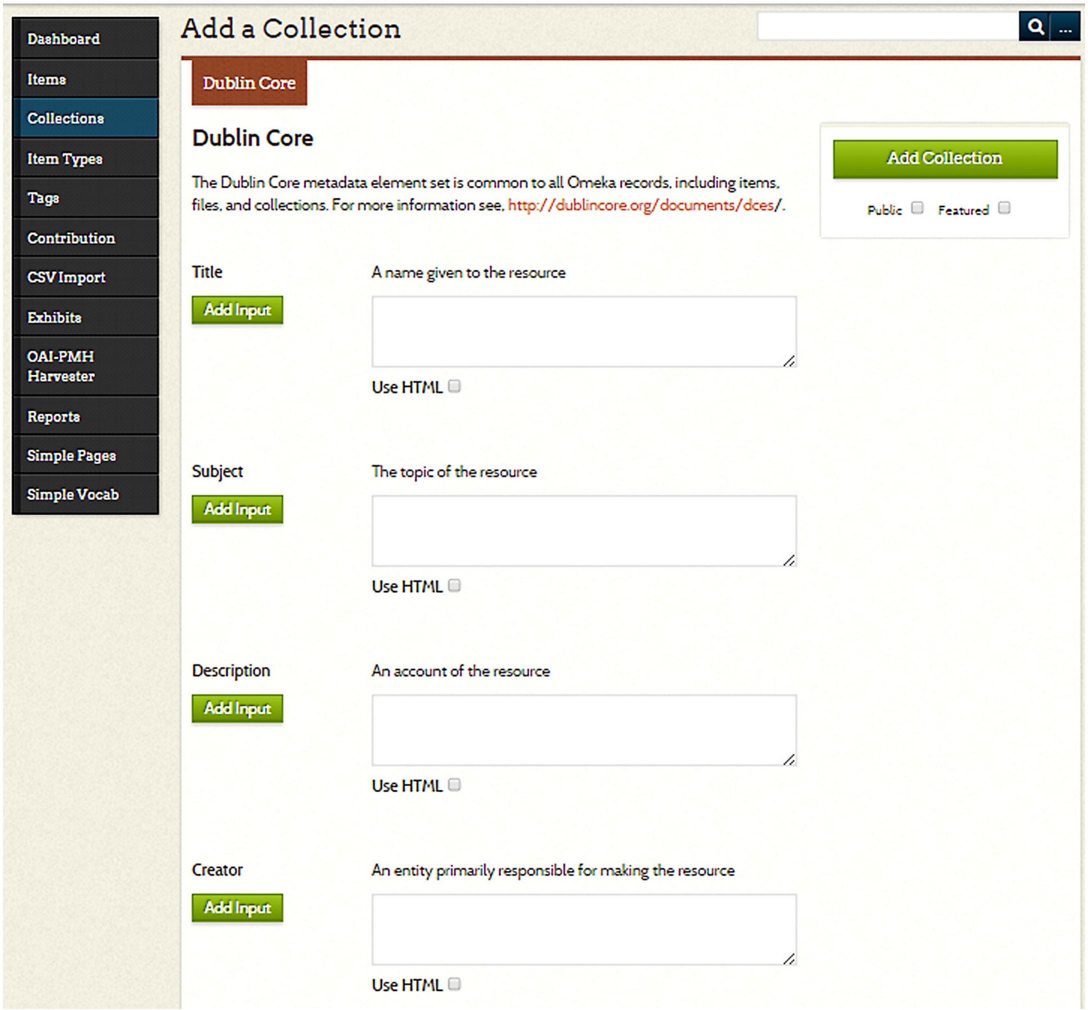
Metadata schemas differ in the type and number of data elements, the designation of mandatory fields, encoding requirements, and the use of data content and value standards. Therefore, a decision about selecting a schema has implications for the quality and level of description. Several factors need to be considered in adopting a schema or developing a customized metadata profile (Kennedy, 2008; Miller, 2011; Zeng and Qin, 2008). Kennedy (2008) offers a practical guide to assist professionals in choosing a metadata schema. The guide consists of nine questions focused on (1) potential users and their needs, (2) expertise of cataloging staff, (3) time and financial resources, (4) type of access to a digital collection, (5) relationships to other collections, (6) collection scope, (7) metadata harvesting, (8) interoperability, and (9) level of maintenance and quality control. Zeng and Qin (2008) point out that the metadata creation process begins with an examination of the discipline, community, and potential users and usage and then considers a number of other criteria, including the nature of the collection and constraints in staffing and funding, as well as institutional and cooperative information systems. Miller (2011) recommends adopting an approach from information architecture that is frequently used to determine functional requirements. The triad of *context, content, and users* can be used as a framework for analyzing the organizational context in which a digital library is created, for examining the type, format, and subject content of materials, and for gathering information about users, their information-seeking behavior, and intended use.

The analysis of user needs and search behaviors provides a foundation for determining functional requirements for metadata from a user viewpoint. Metadata designed with a specific user group in mind allows for not only establishing a specific set of metadata elements but also determining the level of description and selecting appropriate vocabulary. For example, a set of metadata elements for a collection of anatomical images intended for medical students and faculty will be different from a digital collection of anatomy intended for middle school students. Children represent a special user group of digital libraries, and as Abbas (2005) demonstrates, they can benefit from metadata schemas and records developed with their unique needs in mind.

In the process of schema selection and customization, the nature of the collection, the characteristics of the resources, format and subject coverage, user needs, and anticipated use make up the major criteria for consideration. Specific types of materials may require dedicated schemas. As mentioned before, VRA Core and CDWA are schemas developed specifically for creating metadata for works of art. PBCore is often used as a dedicated schema for collections of audio recordings (Dulock and Long, 2011). General digital library schemas, such as Dublin Core and MODS, are used to create item-level descriptive records for a variety of materials, including photographs and monographs, as well as basic records for audio and video resources. Compound objects, such as monographs or newspapers, require structural metadata in addition to descriptive records. As discussed earlier in this chapter, METS is a multifunctional standard capable of providing structural metadata for compound objects as well as serving as a "wrapper" for other types of metadata, including descriptive and preservation metadata.

Digital library management systems (DLMS) are used to build digital collections and repositories. They work as a constraining factor for metadata, as they usually support only a limited number of schemas and controlled vocabulary tools. Chapter 6 reviews a current selection of open source and proprietary DLMS used for constructing digital collections of cultural heritage materials. Repository platforms that primarily serve preservation functions are discussed in Chapter 9. As indicated in the comparative review of DLMS in Chapter 6, system support for metadata schemas varies. A number of DLMS, including CollectiveAccess, CONTENTdm, and EMu provide a selection of schemas and enable customization of metadata templates, while Greenstone, Omeka, and Luna support only Dublin Core. The open source software Omeka is intended for a broad range of developers, including students, scholars, and individuals interested in building personal digital collections. The metadata template includes the basic 15 Dublin Core elements, which cannot be customized (Fig. 5.3). While this approach offers individuals with minimal technical and cataloging expertise an opportunity to build standard-compliant collections, library professionals often find the lack of template customizations limiting (Kucsma et al., 2010).

In the category of open source software, CollectiveAccess provides the strongest support for metadata creation and includes a number of schemas and controlled vocabulary tools. CollectiveAccess also allows users to import and share a variety of standards from user-contributed installation profiles and provides a forum for sharing best practices in metadata creation. Support for metadata design and customization is one of the strengths of proprietary software. CONTENTdm, managed by OCLC and widely adopted by academic and public libraries, offers a number of schemas, including Dublin Core, VRA Core, EAD, and METS. Metadata templates can be customized with local elements mapped to a standard schema.



Add a Collection

Dublin Core

The Dublin Core metadata element set is common to all Omeka records, including items, files, and collections. For more information see, http://dublincore.org/documents/dces/.

Title — A name given to the resource

Subject — The topic of the resource

Description — An account of the resource

Creator — An entity primarily responsible for making the resource

FIGURE 5.3 Selected Basic Dublin Core Elements in Omeka Metadata Template

CONTENTdm also incorporates a number of controlled vocabulary tools, including AAT, TGM, TGN, ULAN, and MeSH. The level of support for metadata creation plays an important role in choosing between an open source software or a proprietary system. Some institutions may use more than one standardized schema for different types of collections if they have access to a DLMS that offers multiple schemas. For example, Dublin Core can be used for image or text collections and EAD for finding aids.

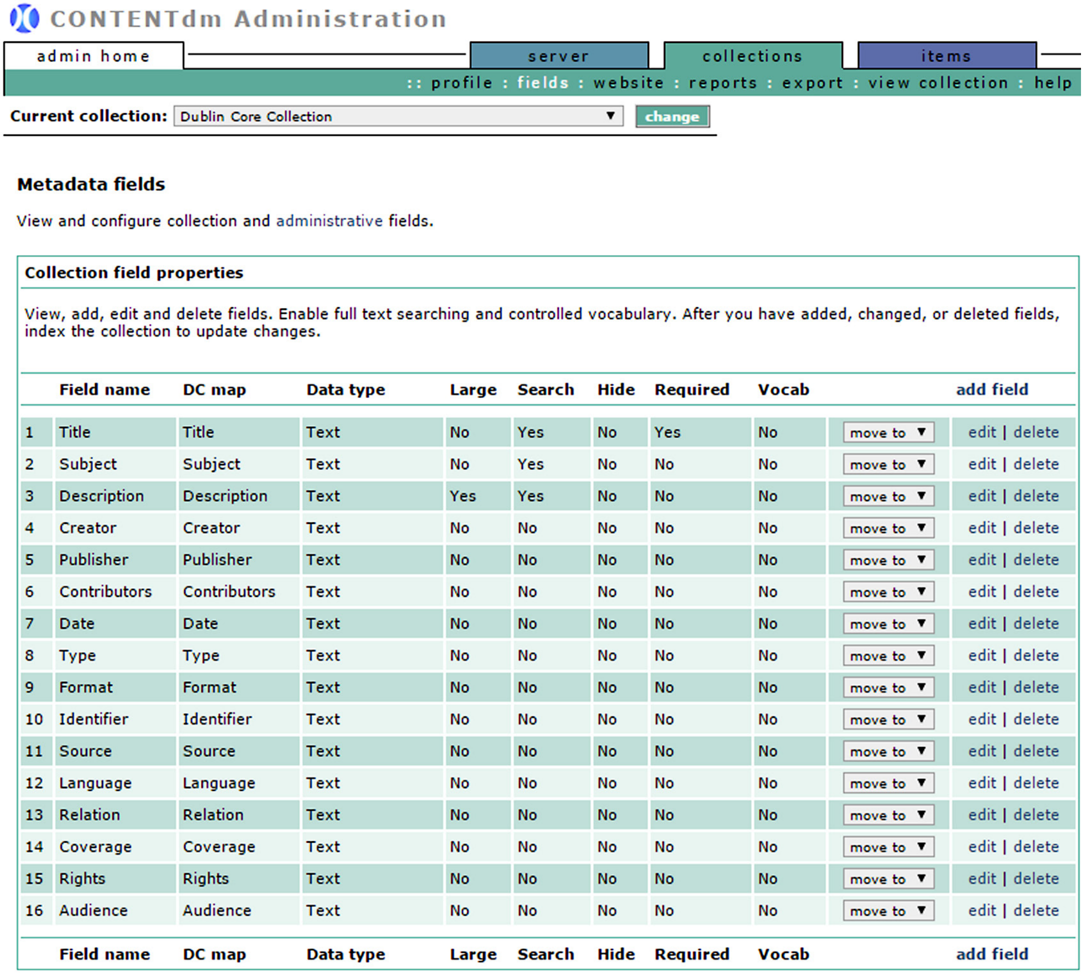
METADATA APPLICATION PROFILES

Analyzing functional requirements and selecting a schema are the first steps in metadata design. Once a standardized schema is selected, metadata designers need to identify which elements have to be included in a collection-specific set. Designing a local metadata application profile should be made in light of a prior analysis of user needs, content characteristics, and organizational context. Standard sets of elements of established schemas, such as Dublin Core, MODS, or VRA Core are often applied without modification if they meet functional requirements or if a content management system (DLMS) lacks capabilities for metadata customization. Chopey (2005) describes many benefits of adopting an established element set for cross-collection searching, metadata sharing, and integration with collections at other institutions. In addition, metadata professionals can follow a user's manual that includes recommended best practices for data values, encoding, and guidelines for data input.

Furthermore, decisions need to be made concerning which elements in a set are designated as required, optional, and/or repeatable. Since flexibility and modularity represent fundamental principles of metadata, elements can be repeated or deemed optional (Duval et al., 2002). Extensibility is another principle of metadata that is supported by a number of schemas, including Dublin Core. Basic elements, such as Date, Description, Coverage, Format, and Relation can be extended and qualified to allow for a greater level of granularity in describing resources. For example, Description can be defined as Description-Abstract or Description-Table of Contents; Coverage as Coverage-Spatial or Coverage-Temporal; Format as Format-Extent and Format-Medium. Date and Relation elements have an extensive list of refinements. In local implementations of schemas, metadata designers can also suppress (hide) selected elements from public viewing, define elements as searchable, and select controlled vocabularies for designated elements. Fig. 5.4 provides an example of the implementation of the basic Dublin Core schema in CONTENTdm. In addition to 15 basic Dublin Core elements, CONTENTdm includes Audience. The software offers collection administrators several options for further refinement of metadata elements.

Miller (2011) emphasizes that designing a good metadata application profile is "dependent on a solid understanding of the meaning and intended scope of the underlying metadata element set standard, such as Dublin Core or MODS; the value and use of controlled vocabularies; and issues of interoperability" (p. 252). Metadata designers need to be familiar with the meaning and usage of elements in the adopted schema in order to implement them correctly in a particular collection and to apply the schema consistently across multiple collections. A number of guides to best practices have been developed to assist digital library practitioners in metadata creation. The Colorado Digitization Program, later known as the Collaborative Digitization Program (CDP), provided one of the first guides to Dublin Core metadata best practices (CDP Metadata Working Group, 2006). The CDP guide has been widely adopted and provides a foundation for the development of many regional and institutional guidelines. Foulonneau and Riley (2008) offer a list of selected local metadata usage guidelines. Chopey (2005) outlines the multiple areas of expertise required for metadata creation in digital collections and repositories, and supplies a checklist for planning and implementing a local metadata application.

Cultural heritage institutions often find the basic set of metadata elements of established schemas insufficient or too restrictive and decide to develop local application profiles (Chopey, 2005). Customized metadata profiles with local elements are usually designed to address particular user needs, disciplinary domains, and characteristics of specific collections. Customized approaches provide more robust or modular metadata structures. Local sets of metadata elements are often designed for heterogeneous digital collections that include materials from multiple source collections or a mix of resource types and formats. However, the challenge of the customized approach is to accommodate and preserve a variety of discipline- or collection-specific metadata, while maintaining consistency across collections and metadata sharing (Attig et al., 2004; Chopey, 2005).


CONTENTdm Administration
admin home
server
collections
items
:: profile : fields : website : reports : export : view collection : help
change

Current collection: Dublin Core Collection ▼

Metadata fields

View and configure collection and administrative fields.

Collection field properties

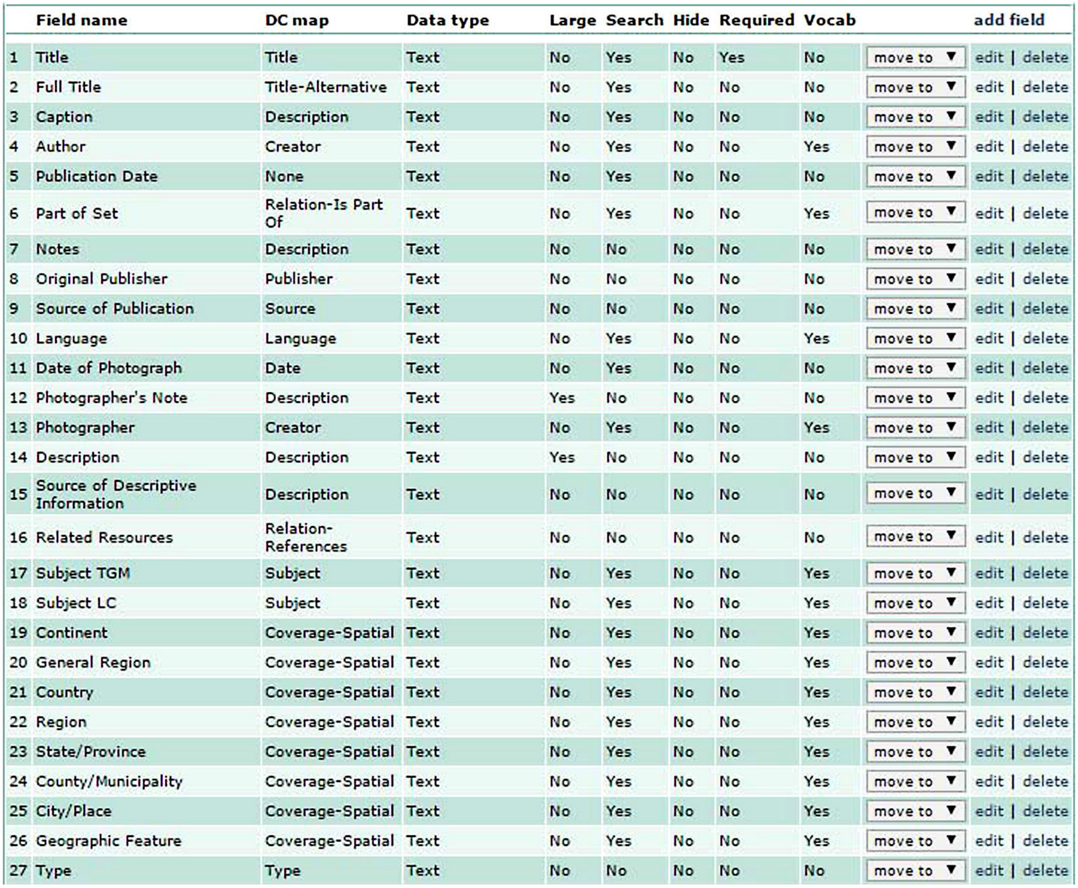
View, add, edit and delete fields. Enable full text searching and controlled vocabulary. After you have added, changed, or deleted fields, index the collection to update changes.

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field
1	Title	Title	Text	No	Yes	No	Yes	No	move to ▼	edit | delete
2	Subject	Subject	Text	No	Yes	No	No	No	move to ▼	edit | delete
3	Description	Description	Text	Yes	Yes	No	No	No	move to ▼	edit | delete
4	Creator	Creator	Text	No	No	No	No	No	move to ▼	edit | delete
5	Publisher	Publisher	Text	No	No	No	No	No	move to ▼	edit | delete
6	Contributors	Contributors	Text	No	No	No	No	No	move to ▼	edit | delete
7	Date	Date	Text	No	No	No	No	No	move to ▼	edit | delete
8	Type	Type	Text	No	No	No	No	No	move to ▼	edit | delete
9	Format	Format	Text	No	No	No	No	No	move to ▼	edit | delete
10	Identifier	Identifier	Text	No	No	No	No	No	move to ▼	edit | delete
11	Source	Source	Text	No	No	No	No	No	move to ▼	edit | delete
12	Language	Language	Text	No	No	No	No	No	move to ▼	edit | delete
13	Relation	Relation	Text	No	No	No	No	No	move to ▼	edit | delete
14	Coverage	Coverage	Text	No	No	No	No	No	move to ▼	edit | delete
15	Rights	Rights	Text	No	No	No	No	No	move to ▼	edit | delete
16	Audience	Audience	Text	No	No	No	No	No	move to ▼	edit | delete
	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field

FIGURE 5.4 Basic Dublin Core Metadata Template in CONTENTdm

The process of developing a customized metadata application profile begins with adopting an existing standardized schema, such as Dublin Core, MODS, and VRA. The established schema is used as the basis for developing a local application profile where some standard elements are retained and some are extended and refined. When a new element is added, it is usually mapped to the standard schema element to enable cross-collection searching and metadata harvesting. Fig. 5.5 demonstrates an example of a customized metadata template in CONTENTdm.

Fig. 5.5 presents a portion of the customized metadata template, which in its entirety consists of 38 elements. This metadata template was designed for a large collection of digitized historic photographs at the American Geographical Society Library, where images were derived from multiple source collections (Matusiak and Johnston, 2014). A basic Dublin Core metadata in CONTENTdm is extended for this project to include a number of refined elements, such as Relation-Is Part Of and



	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab		add field
1	Title	Title	Text	No	Yes	No	Yes	No	move to ▼	edit | delete
2	Full Title	Title-Alternative	Text	No	Yes	No	No	No	move to ▼	edit | delete
3	Caption	Description	Text	No	Yes	No	No	No	move to ▼	edit | delete
4	Author	Creator	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
5	Publication Date	None	Text	No	Yes	No	No	No	move to ▼	edit | delete
6	Part of Set	Relation-Is Part Of	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
7	Notes	Description	Text	No	No	No	No	No	move to ▼	edit | delete
8	Original Publisher	Publisher	Text	No	No	No	No	No	move to ▼	edit | delete
9	Source of Publication	Source	Text	No	No	No	No	No	move to ▼	edit | delete
10	Language	Language	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
11	Date of Photograph	Date	Text	No	Yes	No	No	No	move to ▼	edit | delete
12	Photographer's Note	Description	Text	Yes	No	No	No	No	move to ▼	edit | delete
13	Photographer	Creator	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
14	Description	Description	Text	Yes	No	No	No	No	move to ▼	edit | delete
15	Source of Descriptive Information	Description	Text	No	No	No	No	No	move to ▼	edit | delete
16	Related Resources	Relation-References	Text	No	No	No	No	No	move to ▼	edit | delete
17	Subject TGM	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
18	Subject LC	Subject	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
19	Continent	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
20	General Region	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
21	Country	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
22	Region	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
23	State/Province	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
24	County/Municipality	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
25	City/Place	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
26	Geographic Feature	Coverage-Spatial	Text	No	Yes	No	No	Yes	move to ▼	edit | delete
27	Type	Type	Text	No	No	No	No	No	move to ▼	edit | delete

FIGURE 5.5 Selected Elements in the Customized Dublin Core Template in CONTENTdm

Coverage-Spatial. The elements are often repeated to provide more granular description, especially of geographic coverage. The description of geographic coverage requires special attention throughout the metadata creation process because of the nature of AGS Library's photographic collections, which document geographic expeditions and scientific exploration around the world. The elements related to geographic location include Continent, Subcontinent, Country, Region, Province/State, City/Place, and Geographic Feature. Natural language labels, such as Photographer and Photographer's Note are assigned to accommodate the unique nature of the collection, but all customized elements are mapped to Dublin Core elements, as indicated in the DC column.

Metadata mapping is crucial for searching across collections within a repository and for metadata harvesting to share records in aggregated environments. Elements in local application profiles need to be mapped to semantically corresponding elements in the standard schema, in order to be exposed and shared through the harvesting process. If more than one schema is involved, then interoperability is facilitated by a crosswalk, which maps elements, semantics, and syntax from one metadata schema to those of another (NISO, 2004; Woodley, 2008). Metadata designers may decide not to map some local elements if mapping creates confusion for metadata harvesting.

Adopting an established schema and modifying it to meet collection- or domain-specific requirements are two approaches for creating local application profiles. Zeng and Qin (2008) provide an overview of other models, including profiles that assemble elements from more than one schema. Most of the customized approaches retain interoperability with the original schema base through element mapping and crosswalks. Development of a local element set without reference to an existing standard is generally not recommended, as it locks metadata into a local system without opportunities for sharing and reuse. In the cultural heritage community, it violates a fundamental principle of making metadata sharable and interoperable (NISO Framework Working Group, 2007).

Finally, it is recommended to document the metadata customization process. As Miller (2011) points out, a variety of terms are used in practice for metadata schema documentation, including metadata guidelines, data dictionaries, and metadata application profiles (MAPs). The terms are used interchangeably in the context of digital libraries and usually refer to a document that defines metadata elements for each collection, as well as guidelines for implementation. A data dictionary or metadata application profile provides a list of elements for a given collection, mapping to a standard schema, data format, information about data value tools (authority files and vocabularies), and content guidelines. Many institutions create internal documents, while others choose to share their best practices. University libraries at the University of Washington provide access to their data dictionaries, a.k.a. Schemas and Metadata Application Profiles (or MAPS), at: http://www.lib.washington.edu/msd/pubcat/mig/datadicts.

CONTROLLED VOCABULARIES

Controlled vocabularies/data value standards are essential to the process of standardized metadata creation. While schemas offer a structural framework for building records, controlled vocabularies are a source of authoritative terms to be entered for values of certain elements, such as personal, family, or corporate names, subjects, and coverage elements. The use of controlled vocabularies ensures consistent description of resources and their attributes and enables effective information retrieval and resource discovery. Controlled vocabularies allow the identification of relationships and bring together resources created by the same person or about the same topic. The selection of vocabularies is usually determined during the metadata design and customization process and may involve more than one established controlled vocabulary tool and/or the development of local controlled vocabulary lists.

The use of controlled vocabulary systems is part of a long tradition of bibliographic description in the library world. Digital libraries have adopted the fundamental principles of authority control as well as many tools from the print environment. A controlled vocabulary is defined as "a list or database of subject terms in which all terms or phrases representing a concept are brought together. Often one of the terms or phrases is designated as the preferred term or authorized phrase to be used in metadata records in a retrieval tool" (Taylor and Joudrey, 2008, p. 334). In addition to subject terms, controlled vocabularies can include names of persons, bodies, places, objects, events, and terms for resource type, genre, and format. The term covers a wide range of tools for organizing information retrieval, but at minimum, a controlled vocabulary contains a restricted list of terms. If a metadata element is designed as controlled, only terms from the designated list may be used for entry in metadata records (Hedden, 2008).

Controlled vocabularies address ambiguities and synonymous relationships of natural language at different levels of semantic control. Three types of term relationships are identified in the controlled vocabulary (Leise, 2008):

- Equivalent that includes relations between synonyms and near synonyms—for example, railroads, trains, railways
- Hierarchical that includes relations between broader and narrower concepts—for example, transportation is a broader term (BT) in the hierarchy for railroads, while Cable railroads or Electric railroads represent narrower terms (NT)
- Associative that includes relations of terms that are conceptually related—for example, railroad tracks or railroad bridges

A distinction is made among types of controlled vocabularies because of different levels of semantic relationships (ANSI/NISO, 2005; Leise, 2008):

- Simple controlled lists of terms without any semantic relationships, such as lists of language and country codes, resource type terms, etc.
- Synonym rings that list synonymous terms
- Authority files that list synonyms but also identify a single term as the preferred term, clarifying the equivalency relationship
- Taxonomies that consist of preferred terms connected through hierarchical relationships
- Thesauri that express all three semantic relationships of equivalency, hierarchy, and association

In the context of digital libraries, all five types of controlled vocabularies can be implemented. Typically, thesauri and authority files, including subject heading lists, are used in accordance with the conventions of bibliographic description established in the cultural heritage communities. Subject heading lists are types of authority files that may also include hierarchical or associative relationships. The Library of Congress Subject Headings (LCSH) represent the most widely adopted list of terms for subject description.

Many of the tools used to assign controlled vocabularies in the digital library environment have been developed in the library and museum communities. The Library of Congress offers a number of tools, including:

- *Library of Congress Name Authority File (LC NAF)* as a source of authoritative data for names of persons, organizations, events, places, and titles. Available at: http://id.loc.gov/authorities/names.html
- *Library of Congress Subject Headings (LCSH)* that provide an extensive list of authoritative terms to cover almost all domains of human knowledge. According to Lopatin (2010), LCSH were used for subject terms in digital projects by 87% of academic libraries and 69% of nonacademic libraries.
- *Library of Congress Thesaurus for Graphic Materials (LC TGM)*, which is one of the major thesauri for indexing visual materials. It serves as a source of vocabulary for topical terms in general subject areas and outlines all the semantic relationships among the terms. Available at: http://www.loc.gov/rr/print/tgm1/

The Getty Research Institute has developed a number of controlled vocabulary tools for the museum community. The Getty vocabularies contain controlled terminology for art, architecture, decorative arts, and other material culture, as well as archival materials. Two of the tools, the *Art and Architecture Thesaurus (AAT)* and the *Getty Thesaurus of Geographic Names (TGN)*, have also been adopted broadly in the digital library environment:

- *Art and Architecture Thesaurus (AAT)* includes terms, descriptions, and other information for generic concepts related to art, architecture, conservation, archaeology, and other cultural heritage. Available at: http://www.getty.edu/research/tools/vocabularies/aat/index.html

- *The Getty Thesaurus of Geographic Names (TGN)* provides controlled terms and information for current and historical geographic places and physical features. Available at: http://www.getty.edu/research/tools/vocabularies/tgn/index.html

In addition to the Library of Congress tools and Getty vocabularies, many other general and discipline-specific subject headings and thesauri can be applied in the digital library environment, including Medical Subject Headings (MeSH), UNESCO Thesaurus, National Agricultural Thesaurus, or Iconoclass (Iconographic Classification System).

Fig. 5.6 demonstrates an example of a practical implementation of a number of controlled vocabulary tools in a digital project. It depicts a metadata record from the collection at the American Geographical Society Library based on the template presented in Fig. 5.5. The controlled vocabulary for geographic coverage elements, such as Country, State/Province, and City, is selected from the *Getty Thesaurus of Geographic Names* (*TGN*). The two subject elements are designated to capture different concepts and use different controlled vocabulary tools: Subject TGM covers topical subjects and derives terms from the LC TGM, while Subject LC indicates proper names of people and objects depicted in images and uses LCSH.

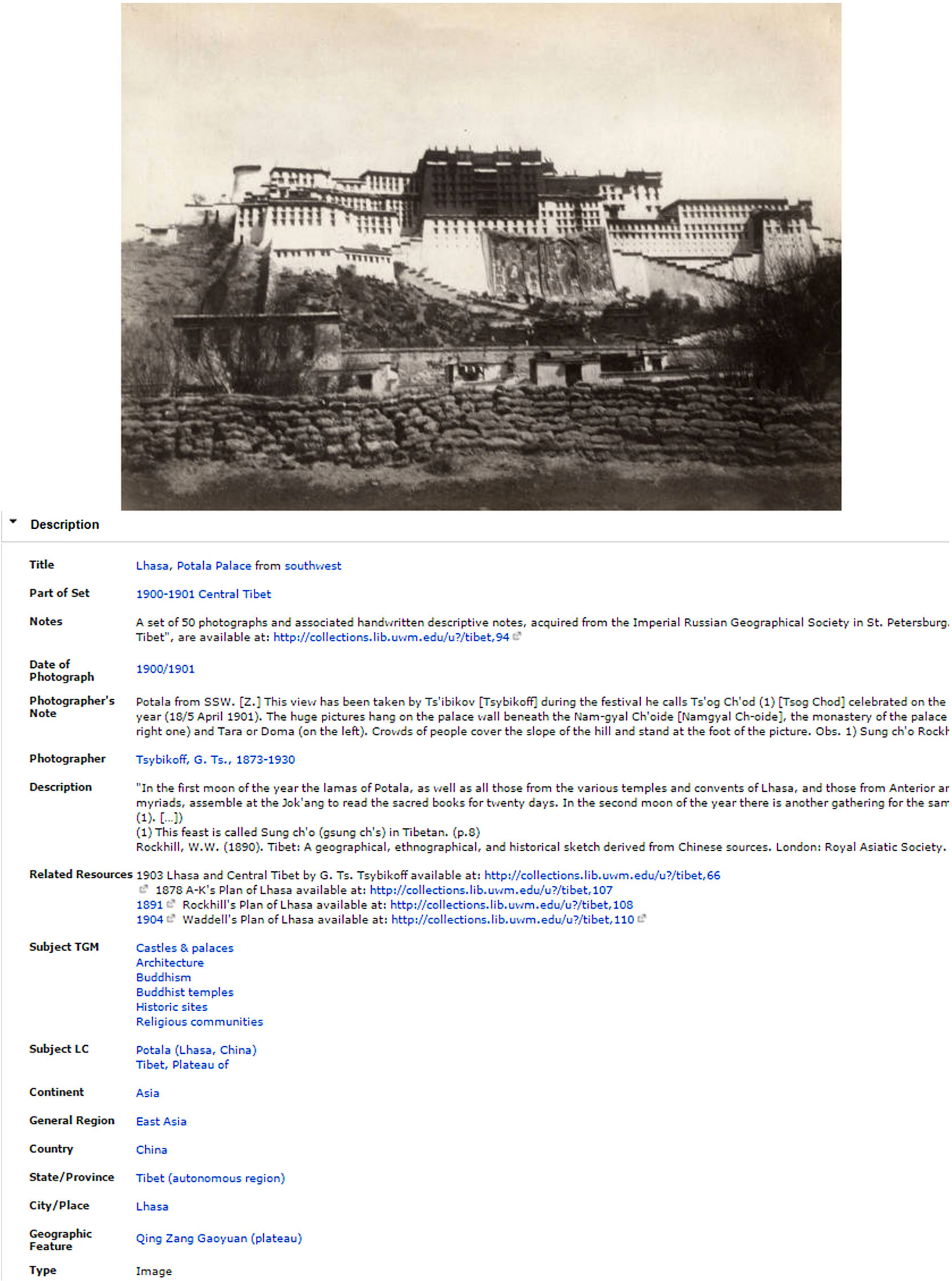
BUILDING METADATA RECORDS

Creating metadata records takes place after a metadata schema has been selected, customized, and documented. Metadata records are encoded in XML, but encoding is usually facilitated by a DLMS that provides templates for building records and generates XML automatically. The process of building records is resource intensive and requires professional staff with knowledge of metadata standards, controlled vocabulary tools, and indexing guidelines. Item-level metadata records need to be constructed for all objects in a digital collection or repository. If original items have limited descriptive information, the process of metadata creation is often accompanied by extensive research to provide accurate descriptions and consistent access points. The process of metadata creation requires:

- Determining resource characteristics
- Transcribing available descriptive information
- Conducting subject analysis
- Selecting appropriate terms from a designated controlled vocabulary tool
- Following content guidelines for data entry
- Recording administrative and preservation information
- Adhering to the established standards

Content guidelines provide directions for the level of description, capitalization, and punctuation. They also specify how to handle variant titles, initial articles, abbreviations, approximate dates, and missing or incomplete information. Metadata designers can adopt an established set of guidelines or develop local guidelines for the purpose of the project or the institutional digital library program. AACR2 general content guidelines have been widely used in bibliographic description. AACR2 is currently being replaced by RDA. DACS and CCO, are domain specific guidelines that are utilized by the archives and museum communities, respectively.

The extent of indexing depends on the type of resource and the amount of descriptive information available in the original collection. As Chopey (2005) notes, creating metadata records for digital



Description

Title	Lhasa, Potala Palace from southwest
Part of Set	1900-1901 Central Tibet
Notes	A set of 50 photographs and associated handwritten descriptive notes, acquired from the Imperial Russian Geographical Society in St. Petersburg. Tibet", are available at: http://collections.lib.uwm.edu/u?/tibet,94
Date of Photograph	1900/1901
Photographer's Note	Potala from SSW. [Z.] This view has been taken by Ts'ibikov [Tsybikoff] during the festival he calls Ts'og Ch'od (1) [Tsog Chod] celebrated on the year (18/5 April 1901). The huge pictures hang on the palace wall beneath the Nam-gyal Ch'oide [Namgyal Ch-oide], the monastery of the palace right one) and Tara or Doma (on the left). Crowds of people cover the slope of the hill and stand at the foot of the picture. Obs. 1) Sung ch'o Rockh
Photographer	Tsybikoff, G. Ts., 1873-1930
Description	"In the first moon of the year the lamas of Potala, as well as all those from the various temples and convents of Lhasa, and those from Anterior ar myriads, assemble at the Jok'ang to read the sacred books for twenty days. In the second moon of the year there is another gathering for the sam (1). [...]) (1) This feast is called Sung ch'o (gsung ch's) in Tibetan. (p.8) Rockhill, W.W. (1890). Tibet: A geographical, ethnographical, and historical sketch derived from Chinese sources. London: Royal Asiatic Society.
Related Resources	1903 Lhasa and Central Tibet by G. Ts. Tsybikoff available at: http://collections.lib.uwm.edu/u?/tibet,66 1878 A-K's Plan of Lhasa available at: http://collections.lib.uwm.edu/u?/tibet,107 1891 Rockhill's Plan of Lhasa available at: http://collections.lib.uwm.edu/u?/tibet,108 1904 Waddell's Plan of Lhasa available at: http://collections.lib.uwm.edu/u?/tibet,110
Subject TGM	Castles & palaces Architecture Buddhism Buddhist temples Historic sites Religious communities
Subject LC	Potala (Lhasa, China) Tibet, Plateau of
Continent	Asia
General Region	East Asia
Country	China
State/Province	Tibet (autonomous region)
City/Place	Lhasa
Geographic Feature	Qing Zang Gaoyuan (plateau)
Type	Image

FIGURE 5.6 Metadata Record from a Collection Built Using a Local Application Profile

The record is available at: http://collections.lib.uwm.edu/cdm/ref/collection/tibet/id/129.

collections requires more granular indexing than the kind of bibliographic description found in library catalogs. Archival image collections are particularly challenging because very few items will have individual annotations, and the level of consistency and accuracy of description may vary from item to item. On the other hand, monographs usually have MARC cataloging records, so the metadata process can be streamlined. In these cases, MARC-Dublin Core or MARC-MODS crosswalks can be used to automate metadata creation. An item-level metadata record details the characteristics of a digital object for the purposes of description, resource discovery, and preservation. It typically includes:

- Descriptive information
- Access points
- Contextual information
- Reference to the original item and collection
- Administrative and preservation information

Finally, metadata records need to be reviewed for quality and consistency. Ultimately, the quality of metadata and adherence to standards determine if digital objects are findable and discoverable within local digital collections, as well as in the aggregated environment of large-scale digital libraries.

USER TAGGING

The emergence of Web 2.0 technologies has challenged the traditional approaches to description and organization of digital library materials and offered new opportunities for user engagement and knowledge contribution (Alemu et al., 2012; Matusiak, 2006; Trant, 2009). Web 2.0 emerged in 2004 and transformed the web from a static platform into a dynamic, shared information space (Ding et al., 2009). In contrast to Web 1.0, the network of hyperlinked but relatively static documents, Web 2.0 introduced a participatory and interactive model where users can contribute and actively engage with web content. Web 2.0 encompasses a wide range of web applications that enable users to share their own resources and comments on the content of others.

User tagging is particularly relevant in digital libraries, as it offers an opportunity to enhance metadata created by library professionals by introducing user language and perspective to contribute additional descriptive information (Matusiak, 2006; Trant, 2009). User tagging represents an approach to organizing content in the web environment where users create their own textual descriptors using natural language terms (tags) and share them with a community of users. This new system of organization that employs users to assign keywords to their own or shared content, has been referred to by several terms, including user tagging, social tagging, collaborative tagging, folksonomy, social classification (Hammond et al., 2005), or “metadata for the masses” (Merholz, 2004).

The potential of user-generated descriptive tags for library resources has caught the attention of the researchers and practitioners, resulting in an extensive body of literature devoted to examining the benefits of tagging and comparing tags to structured metadata (Bar-Ilan et al., 2008; Kipp, 2011; Matusiak, 2006; Petek, 2012; Pirmann, 2012; Rorissa, 2010). The main benefits include a more user-centered approach to describe resources, closer connection to users and their language, user engagement, and collaborative knowledge construction. Rorissa (2010) indicates that user tags and traditional assigned index terms have different structures. Moreover, user tags reflect users’ context and can be semantically richer than index terms. At the same time, professional indexers use controlled vocabularies and thoroughly evaluate a document to achieve higher precision when users are searching for a resource. In short, user-generated tagging is a double-edged sword. On one hand, tags are criticized for imprecision and inaccuracy; on the other hand, they are able to capture the breadth of user language. After analyzing tagging and controlled vocabulary studies, Thomas et al. (2009) emphasize that user tags can enhance controlled vocabularies by offering additional access points. Kipp (2011) concludes "tagging does not completely replace controlled vocabularies but provides an added dimension to subject access from the perspective of the end users" (p. 30).

Studies comparing user tags and user queries represent another area of research. After analyzing user tagging and user queries, Ransom and Rafferty (2011) confirm that user tagging can help the effectiveness of information retrieval. In particular, the authors find similarities between tags and search terms associated with people, objects, and location. Benoit (2014) further compares expert and novice user tags and investigates how these tags match with query terms. The results reveal that expert tags match query terms more than novice tags, while the combination of expert and novice tags shows the highest matching of query terms. Huang and Jörgensen (2013) investigate differences in tagging between digital collections and social sites, such as Flickr. In general, popular tags in Flickr describe more generic objects, while the popular tags identify more specific objects and time categories in the Library of Congress' photostream (LCP).

Despite the advantages of engaging users with digital library resources through tagging, the applications of user-generated tags have not been widely implemented in digital libraries. According to a 2010 survey, only 9% of the academic libraries and 25% nonacademic libraries enable user-generated metadata (Lopatin, 2010). A variety of approaches have been applied to engage users with digital libraries. For example, Bainbridge et al. (2012) designed a client-facing JavaScript browser extension that allows users to edit, merge, delete, and undo metadata elements in digital libraries.

Researchers and practitioners also identify a range of challenges with user tagging (Guy and Tonkin, 2006; Macgregor and McCulloch, 2006; Matusiak, 2006; Rorissa, 2010; Thomas et al., 2010). Some of the most common issues include:

- Misspellings or unidentifiable terms
- Imprecise and unclear tags
- Uncontrolled and inconsistent tags (e.g., variations of the same tags)
- Lack of authority control (e.g., synonyms)
- Increased recall, but low precision
- Lack of collocation

In addition, Jeong (2009) discovers that a high ratio of overlap between tags and metadata elements, such as title and description, reduces the effectiveness of tagging in information organization and retrieval. Lu et al. (2010) point out that some tags are personal and subjective rather than subject related, which hinders the integration of tags into library systems. Bar-Ilan et al. (2008) compare structured and unstructured tagging in a cultural heritage collection, and they find that different interpretations of the meaning of structured elements reduce the quality of tagging.

In addition to identifying the challenges, researchers and practitioners have also offered some suggestions for overcoming the limitations of unstructured and inconsistent tags and integrating them into standardized metadata. Thomas et al. (2010) recommend a number of solutions, including providing users with guidelines for tag creation, enabling users to edit and combine tags, and linking tags to controlled vocabularies. Since both user tags and controlled vocabularies have strengths and weaknesses, researchers propose integrating controlled vocabularies and user tags in digital library systems (Pirmann, 2012; Thomas et al., 2010).

Since their inception, Web 2.0 applications are gradually becoming integrated into DLMS. Both open source and proprietary solutions, such as Omeka, CollectiveAccess, and CONTENTdm offer technical capabilities for engaging users in the tagging of digital objects, contributing comments, and sharing resources through social media. Although many digital library systems support user tagging, in practice user-generated tags remain limited. However, the institutions that expose some of their digital library resources to the general public through social media have had more success with engaging users in tagging. Flickr: The Commons was initiated as a collaborative project between Flickr and the Library of Congress to increase the visibility of cultural heritage materials and to provide a way for the general public to contribute to the description of resources (Clark, 2008). Since the launch in 2008, many other cultural heritage institutions have decided to join Flickr: The Commons to expose their collections and take advantage of user contributions (Flickr, 2015).

LINKED DATA

The concept of linked data is associated with the Semantic Web, also referred to as the web of Data or Web 3.0. The vision of the Semantic Web goes beyond the functionality of Web 1.0 or the social interactions of Web 2.0. It aims to establish a global network of data from diverse domains, connected through semantic relationships that are not only understood by humans but can also be accessed and interpreted by computers (Berners-Lee et al., 2001; W3C, 2015a). Berners-Lee et al. (2001) envision the Semantic Web as "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (p. 28). The ultimate goal is to use computing capabilities to enhance discovery of the related information, share and reuse data in the open web environment, and enrich knowledge through linking data among multiple domains. In order to make the Semantic Web a reality, data need to be open, structured, and connected through a set of standards and technologies that not only process the data but also build meaningful links among different data sets (W3C, 2015b; Yoose and Perkins, 2013). The collection of interrelated datasets on the web, referred to as linked data, is at the heart of the Semantic Web (W3C, 2015b).

Linked data rely on a stack of technologies to establish semantic relationships and publish interrelated data sets, but it is not a specific standard or technology. It is often described as a set of best practices for the publication of structured data on the web (Bizer et al., 2009; Van Hooland and Verborgh, 2014). RDF (Resource Description Framework) provides a conceptual data model for establishing relationships and representing linked information on the web. The underlying practices and technologies for developing linked data or transforming existing metadata into linked data sets are still evolving. It is an emergent field with a growing number of standards and open source tools for encoding, publishing, and retrieving linked data.

Linked data encompass all types of structured data that can be interlinked, published openly on the web, and searched through semantic queries. Increasingly, the concept is gaining attention in the library world because of its potential to address the limitations of the current metadata practices and to move library metadata from its storage silo in library systems and databases to the open web (Coyle, 2012; W3C, 2011). Singer (2009) describes problems with the quality of metadata and the isolation of library information systems in an article advocating for the adoption of linked data:

> We have silo sitting next to silo, with much duplication of data; arcane, inefficient, and sometimes completely broken methods of determining that two records are describing the same thing; and very little control over relating one resource in one system to another in a completely different application (even if these serve a similar purpose), much less data available outside the institution (p. 114).

While Singer's description may appear overly critical, it does point to a fundamental issue that remains largely unresolved despite significant efforts to improve interoperability, federated searching, and metadata harvesting. Library bibliographic and digital collection metadata, stored in separate databases, are poorly connected within the library information landscape. The separation of library systems from the open web represents an even more critical issue. The wealth of library metadata is not easily accessible to search engines requiring users to search individual catalogs, digital collections, and repositories. As the authors of *The Library Linked Data Incubator Group Final Report* note, although library databases do have searchable interfaces, library data are not integrated with web resources (W3C, 2011).

Metadata interoperability standards were developed to address the issues of resource discovery and sharing in the digital library field. The presence of multiple metadata standards and customized approaches, however, hinders interoperability and metadata harvesting in distributed environments. As Van Hooland and Verborgh (2014) point out, even if the institutions adopt metadata standards, they often implement them in a different way to accommodate the specific nature of their collections. Metadata harvesting also results in some information loss when rich metadata records are reduced to basic Dublin Core elements. The disadvantage of federated searching is the lack of granularity and the inability to support advanced queries. Lampert and Southwick (2013) note that aggregated collections "lose the richness of their original metadata when added to systems designed to enhance discovery," and cite this shortcoming as one of the reasons for embracing linked data (p. 236).

Discussions of linked data in library literature usually begin by pointing out the need for a new approach to metadata structuring and outlining the potential benefits of transforming library metadata into linked data (Alemu et al., 2012; Byrne and Goddard, 2010; Coyle, 2012; Lampert and Southwick, 2013; Mitchell, 2013a; Singer, 2009). The focus is primarily on breaking the walls surrounding library resources, exposing rich library metadata, and connecting them to related information on the web. Byrne and Goddard (2010) list a common format for all data in the linked data environment as one of the major benefits that can improve the interoperability and integration of library systems. Significant advantages of the linked data approach over current metadata practices for multiple library stakeholders are outlined in *The Library Linked Data Incubator Group Final Report* (W3C, 2011).

Exposing library metadata via the open web requires fundamental restructuring of data models and a radically different approach to recording metadata (Mitchell, 2013a; Van Hooland and Verborgh, 2014). The current metadata practices in the digital library field rely primarily on relational or XML data models with a central concept of a record governed by a schema. Metadata records, following the legacy of MARC, have a flat structure, in which all metadata statements about an object's properties (title, author, subject, etc.) are contained in a single record. Metadata schemas, such as Dublin Core or MODS, are more flexible and extensible than MARC; nonetheless they remain static, constrained by the concept of a record. Metadata sharing between different collections and domains requires the mapping of metadata elements and the adherence to a common schema. XML represents a significant step toward automatic sharing of data as it provides a standardized syntax for the exchange of structured data (Van Hooland and Verborgh, 2014). The sharing of XML-based metadata in practice, however, can be difficult because of the reliance on schema structures. RDF, the data model underlying linked data, offers greater flexibility since it moves away from a record structure and it doesn't require a schema to interpret and reuse data.

LINKED DATA MODEL AND TECHNOLOGIES

Linked data represent a radical shift in the way structured data can be created to express information about resources. Instead of a record-based model governed by a schema, it focuses on smaller chunks of meaningful metadata that can be linked and queried. In this environment, metadata statements, rather than records, represent a basic unit of metadata. This approach for structuring data is schema-neutral, but it does use a range of standardized vocabularies to define classes and properties of resources and the relationships between them (Van Hooland and Verborgh, 2014). RDF provides a data model for making simple statements and connecting them in a series. The process of developing or transforming existing metadata into linked data is based on a set of principles and relies on a number of technologies and tools. Mitchell (2013b) provides an overview of five building blocks of linked data:

1. RDF data model for structuring statements
2. Content rules
3. RDF-compatible metadata schemas and vocabularies
4. Serialization formats for encoding RDF statements
5. Technologies for publishing and exchanging linked data, including the SPARQL protocol

A full description of linked data technologies is outside of the scope of this chapter. The focus of the following section is primarily on the conceptual aspects of data modeling.

RDF provides a foundation for building linked data. It is an abstract data model used to express and interlink meaningful pieces of information and to represent them on the web. RDF offers a common framework in which information can be exchanged between applications without losing the meaning (Working Group, 2014). RDF statements are constructed as triples and consist of subjects, objects, and predicates. Any resource (subject) can have a relationship (predicate) to another resource (object). Resources can be conceptual, physical, or digital. Objects are used to express descriptions of resources (subjects), while predicates specify how the subjects and objects are related (Fig. 5.7 for a visual representation of an RDF triple). The model is flexible so that objects can become subjects in another series of statements. This simple syntax can be used to capture statements about resources. For example, for a digital image that shows a nomadic woman in Tibet (Fig. 5.8), several statements can be created, as demonstrated in Table 5.4.

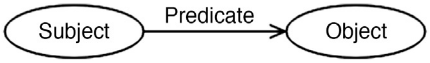


FIGURE 5.7 The Structure of an RDF Triple

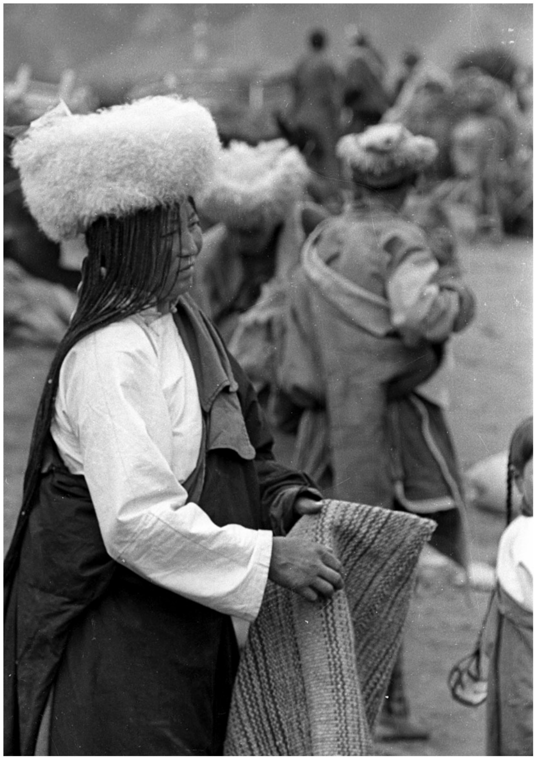


FIGURE 5.8 Nomadic Tibetan Woman with Fur Hat in Tibetan Plateau

The Dublin Core metadata record available at: http://collections.lib.uwm.edu/cdm/ref/collection/tibet/id/1013.

Multiple statements, corresponding to the elements and their values in the metadata record, can be made about this image. The RDF statements also allow a more granular level of description. Another series of statements could be constructed about an original film negative to make a distinction between the analog image (physical object) and its digital representation, thus disambiguating what is being described. Moreover, the flexible nature of the RDF model allows expansion of relationships and their connections to external resources. For example, the photographer Harrison Forman is also the author of a book,

Table 5.4 Statements about a Digital Image of Nomadic Tibetan Woman, Digital ID fr203647

Subject	Predicate	Object
Digital image fr203647	hasTitle	Nomadic Tibetan woman with fur hat in Tibetan Plateau
Digital image fr203647	hasCreator	Harrison Forman
Digital image fr203647	hasSubject	Nomads
Digital image fr203647	hasSubject	Tibet Autonomous Region (China)–Social life and customs

Table 5.5 Statements About a Book, *Through Forbidden Tibet*, LC Control Number 35025394

Subject	Predicate	Object
Book LCCN 35025394	hasTitle	*Through Forbidden Tibet: An Adventure into the Unknown*
Book LCCN 35025394	hasCreator	Harrison Forman
Book LCCN 35025394	hasSubject	Tibet Autonomous Region (China)–Social life and customs

Through Forbidden Tibet: An Adventure into the Unknown. Published in 1935, this book provides an account of Forman's travels through Tibet, as well as useful context for his photographic record. If a series of statements is constructed about the book *Through Forbidden Tibet,* the digital collection metadata and bibliographic data can be connected. Table 5.5 presents a sample of RDF statements about the book.

The digital image fr203647 and the book LCCN 35025394 can be linked through the RDF statements. The two resources can be further interconnected through a subject relationship: "Tibet Autonomous Region (China)—Social life and customs." RDF enables the creation of multiple statements about a resource. RDF triples are connected in a series of statements (serialization). Serialization is the process of expressing RDF triples/statements in a machine-processable syntax such as RDF-XML, Turtle, JSON-LD, etc. The RDF statements in Table 5.4 and Table 5.5 outline a few basic semantic relationships using natural language statements, but the real strength of linked data is in expressing the relationships through Uniform Resource Identifiers (URIs) so that the resources can be linked and queried on the web.

The URI identifies the name and/or location of a file or resource in a uniform format. The use of URIs for the identification of resources, their values, and relationships is part of the core principles of linked data formulated by Tim Berners-Lee (2006):

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs so that they can discover more things

The use of globally shared URIs is a fundamental concept of linked data. The URIs can be used to identify unambiguously any kind of object or concept. URI-based statements facilitate the building of complex relationships using external standards and vocabularies. Various linked open data vocabularies (LOVs) can be used as a source of URIs. It is recommended to reuse existing URIs available through linked data sources if possible (Bizer et al., 2009; Mitchell, 2013b; W3C, 2011). In some cases, URIs are locally assigned, for unique resources or locally controlled vocabulary terms (Lampert and Southwick, 2013; Southwick, 2015).

LOVs represent another major building block in the process of constructing and transforming existing metadata into linked data (Mitchell, 2013b). LOV is an umbrella term that encompasses a range of schemas and ontologies as well as value vocabularies. Ontologies are semantic models of the things, entities, or concepts that exist in a specific knowledge domain. The Web Ontology Language (OWL) is a full-fledged ontology language for developing ontologies and LOVs. The Resource Description Framework Schema (RDFS) is a basic level ontology language for defining relationships. Semantic Web data modeling standards such as OWL and RDFS share some similarities with traditional knowledge organization systems, but also differ from them in several fundamental respects, especially in that they are designed to allow for semantic querying and machine processing (Miller, 2015).

Schemas define classes and properties for linked data, while value vocabularies are a source of URIs for resources and their values. In the context of digital libraries many traditional metadata schemas and controlled vocabularies are being adopted for the linked data environment and transformed following the specifications provided by semantic data models, such as RDFS. A distinction can be made between:

- Metadata element sets published as RDF vocabularies. Several digital library metadata schemas have been defined as RDF vocabularies, with Dublin Core being the most frequently used. The Dublin Core Metadata Initiative (DCMI) has put significant efforts into adopting the Dublin Core Element Set for implementation in the linked data environment (DCMI, 2015b). All Dublin Core terms that conform to the DCMI Abstract Model are assigned a unique URI that provides a vocabulary for expressing relationships in RDF.
- Value vocabularies include authority files, taxonomies, subject headings, thesauri, and classification systems that have assigned unique URIs to their entries. A number of controlled vocabularies maintained by the Library of Congress and the Getty Research Institute have been made available in the linked data format (Library of Congress, 2015d; Getty Research Institute, 2015). Another example of a published linked data vocabulary is Virtual International Authority File (VIAF), developed as a collaborative project of several national libraries. DBpedia, one of the largest repositories of linked data vocabulary, was created by extracting structured information from Wikipedia (Dbpedia, 2015). Several cultural heritage institutions use Dbpedia as a source of vocabulary in their linked data projects (Pattuelli and Rubinow, 2013; Southwick, 2015).

As is the case with standard schemas and controlled vocabularies, a variety of linked open vocabularies are available, requiring the selection of appropriate tools during the implementation process. The example explored in this chapter demonstrates how several LOVs are used to assign URIs to statements. Table 5.6 shows several sets of RDF statements constructed for a digital image fr203647 and expressed as URIs, while Table 5.7 presents a set of triples for the book.

Table 5.6 Statements for a Digital Image fr203647 Expressed as URIs

Subject	Predicate	Object
http://collections.lib.uwm.edu/ProvidedCHO/fr203647[a]	http://purl.org/dc/elements/1.1/title	Nomadic Tibetan woman with fur hat in Tibetan Plateau
http://collections.lib.uwm.edu/ProvidedCHO/fr203647	http://purl.org/dc/elements/1.1/creator	http://id.loc.gov/authorities/names/n88172344
http://collections.lib.uwm.edu/ProvidedCHO/fr203647	http://purl.org/dc/elements/1.1/subject	http://id.loc.gov/vocabulary/graphicMaterials/tgm007097
http://collections.lib.uwm.edu/ProvidedCHO/fr203647	http://purl.org/dc/elements/1.1/subject	http://id.loc.gov/authorities/subjects/sh2008117270

[a] *The URI for the subject (digital image fr203647) is fictional since this object in the UWM Digital Collections does not have a persistent URI. The reference URL available for the object <http://collections.lib.uwm.edu/cdm/ref/collection/tibet/id/1013> is generated by CONTENTdm and is software dependent.*

Table 5.7 Statements about a Book, *Through Forbidden Tibet*, Expressed as URIs

Subject	Predicate	Object
http://lccn.loc.gov/35025394	http://purl.org/dc/elements/1.1/title	*Through Forbidden Tibet: An Adventure into the Unknown*
http://lccn.loc.gov/35025394	http://purl.org/dc/elements/1.1/creator	http://id.loc.gov/authorities/names/n88172344
http://lccn.loc.gov/35025394	http://purl.org/dc/elements/1.1/subject	http://id.loc.gov/authorities/subjects/sh2008117270

The predicate terms (title, creator, and subject) have URIs assigned from the Dublin Core RDF-compatible vocabulary. The Library of Congress linked data sets: LC Authority File, Thesaurus for Graphic Materials, and LC Subject Headings are sources of URIs for creator and subject values. Title is the only element that has literal value because it is unique and does not belong to a controlled vocabulary.

Table 5.7 shows a sample of triples for Forman's book. The URI to represent this book uniquely (subject) is assigned by following the Library of Congress permalink. Again, title is the only element that has a literal value. The use of URIs for identifying resources uniquely, expressing relationships, and recording values, represents a significant departure from the digital library practices where so far metadata has been recorded as natural language descriptions and controlled vocabulary terms encoded as text.

The RDF data model and URIs provide a foundation for creating semantic relationships and constructing unambiguous links. An additional set of tools or "building blocks" is needed to encode and publish linked data sets so they can be processed by computers and rendered in formats usable and accessible to end users. Ultimately, linked data sets need to be presented through interfaces supporting semantic queries. RDF statements have to be encoded in a machine-readable syntax or serialization format in order to be stored and queried. It is beyond the scope of this chapter to review the rather complex stack of linked data technologies, but it is worth mentioning that several serialization formats are currently available. Mitchell (2013b) provides an overview of commonly used formats, including RDF/XML, RDF Notation-3/N3, Turtle, RDFa, and JSON-LD. Finally, a variety of tools is used to support the storage and exchange of linked data. SPARQL (SPARQL Protocol and RDF Query Language) is a W3C recommendation that provides a set of specifications to govern the query structure and a protocol for querying and exchanging data (Mitchell, 2013b).

LINKED DATA AND DIGITAL LIBRARIES

Linked data represent an emergent but rapidly growing area in digital library research and practice, with innovative and collaborative projects in the cultural heritage community. The emphasis of digital library efforts is on open data free of copyright restrictions with the term "linked open data" (LOD) frequently used in the library, archives, and museum (LAM) community. Although linked data technically do not need to be open in order to be interoperable, opening data increases the potential of linked data technology and makes data sharable and reusable (W3C, 2011). Opening data means providing the data freely, without copyright or other rights restrictions. LODLAM is an acronym for Linked Open

Data in Libraries, Archives, and Museums that refers to an informal network of scholars and practitioners engaged in the research and implementation of linked data technology in digital collections and repositories (LODLAM, 2015). The focus of digital library research and practice activities is on transforming library metadata into LOD and developing LOV.

In the context of digital libraries, adopting linked data requires a transformation of the existing schemas, controlled vocabulary tools, and record-based metadata sets into linked data formats. Although linked data represents a new approach to data modeling and recording metadata, it also builds upon the existing digital library schemas and vocabularies (Alemu et al., 2012; Yoose and Perkins, 2013). The foundational *Library Linked Data Incubator Group Final Report* (W3C, 2011) provides a set of recommendations for moving forward with the process of transforming library metadata into linked data. The key recommendations are:

- Identifying sets of data as possible candidates for early exposure as linked data and fostering a discussion about open data and rights
- Increasing library participation in Semantic Web standardization and developing library data standards that are compatible with linked data
- Creating URIs for the items in library datasets, developing policies for managing RDF vocabularies and their URIs, and expressing library data by reusing or mapping to existing linked data vocabularies
- Preserving linked data element sets and value vocabularies and applying library experience in the curating and long-term preservation of linked datasets

The authors of the Report recommend an incremental approach, noting that an effort to expose the complexity of library data as linked data all at once could be disruptive and "have limited success" (W3C, 2011, Section 4.1.1). However, some library tools, such as authority files, subject headings, and thesauri, lend themselves easily to publication as linked data. As mentioned before, several controlled vocabulary tools maintained by the Library of Congress and the Getty Research Institute have been released as LOVs in recent years (Library of Congress, 2015d; Getty Research Institute, 2015).

The transformation of metadata records in digital collections and repositories into linked data sets represents a major undertaking. Again, the process has been moving gradually from prototypes and research experiments into practical implementations. Linked data projects range from the national and large-scale digital library initiatives to smaller efforts undertaken by individual cultural heritage institutions. The number of publications and case studies documenting the process and sharing lessons in linked data development, although still limited, is growing (Hatop, 2013; Lampert and Southwick, 2013; Mitchell, 2013c; Pattuelli and Rubinow, 2013; Pattuelli et al., 2013; Southwick, 2015). Van Hooland and Verborgh (2014) provide a number of case studies in their book *Linked Data for Libraries, Archives and Museums*, while Yoose and Perkins (2013) review major LOD projects and initiatives undertaken in the library, archives, and museum communities, including SNAC (Social Networks and Archival Context Project), LOCAH (the Linked Open Copac and Archive Hub Project), and a linked data project at the Smithsonian American Art Museum. Europeana and the Digital Public Library of America (DPLA), two large-scale digital libraries and metadata aggregation platforms, are actively engaged in linked data by promoting open metadata and providing a range of linked open data resources and services (Mitchell, 2013c). Europeana and the DPLA are discussed in more detail in Chapter 1 and Chapter 11.

Several researchers stress that libraries are uniquely positioned to adopt linked data because of a strong tradition of standardization, the use of controlled vocabularies, and some experience in interoperability (Bair, 2013; Byrne and Goddard, 2010; Coyle, 2012). On the other hand, the use of library-specific standards, the disparity between the library and Semantic Web terminology, the lack of unique URIs for most library resources, and finally the complexity of linked data technology pose significant obstacles to a widespread adoption of linked data in the library world (Alemu et al., 2012; Byrne and Goddard, 2010; W3C, 2011). Byrne and Goddard (2010) note that most of the barriers are of a non-technical nature and identify the lack of awareness as a fundamental challenge for the development of linked data in libraries.

Metadata in digital libraries seems to be at a crossroads after two decades of intensive standardization and development. Linked open data offer an opportunity to integrate digital library objects with other library resources and make them more visible on the web but also requires a significant restructuring of existing metadata sets. The two models of record-based metadata and RDF-modeled linked data may coexist for a while in the digital library universe, but the current metadata practices can also be adopted in preparation for moving metadata into linked open data formats with closer attention to metadata quality, schema mapping, and a standardized assignment of unique URIs to digital library objects. As awareness in the professional community and practical experience in developing LOD increase, the body of linked metadata sets will grow, transcending the barriers between the current digital library systems and the open web.

REFERENCES

Abbas, J., 2005. Creating metadata for children's resources: issues, research, and current developments. Lib. Trends 54 (2), 303–317.

Alemu, G., Stevens, B., Ross, P., 2012. Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: a social constructivist approach. New Lib. World 113 (112), 38–54.

Alemu, G., Stevens, B., Ross, P., Chandler, J., 2012. Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models. New Lib. World 113 (11/12), 549–570.

ANSI/NISO, 2005. American National Standards Organization (ANSI) and National Information Standards Organization (NISO). ANSI/NISO Z39.19-2005 (R2010). Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Available from: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf.

Arms, W.Y., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J., Van de Sompel, H., 2002. A spectrum of interoperability: the site for science prototype for the NSDL. D-Lib Magazine, 8 (1), 9. Available from: http://www.dlib.org/dlib/january02/arms/01arms.html.

Attig, J., Copeland, A., Pelikan, M., 2004. Context and meaning: the challenges of metadata for a digital image library within the university. Coll. Res. Lib. 65 (3), 251–261.

Baca, M., Harpring, P., 2009; revised 2014. Categories for the Description of Works of Art. J. Paul Getty Trust College Art Association. Available from: http://www.getty.edu/research/publications/electronic_publications/cdwa/index.html.

Baca, M., 2008. Glossary. In: Baca, M. (Ed.), Introduction to Metadata. The Getty Research Institute, Los Angeles (CA), pp. 73–78.

Bainbridge, D., Twidale, M.B., Nichols, D.M., 2012. Interactive context-aware user-driven metadata correction in digital libraries. Int. J. Digital Lib. 13 (1), 17–32.

Bair, S., 2013. Linked data—the right time? J. Lib. Metadata 13 (2–3), 75–79.

Barbero, G., Trasselli, F., 2014. Manus OnLine and the Text Encoding Initiative Schema. J. Text Encoding Init., (8-Preview). Available from: http://jtei.revues.org/1054.

Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., Shachak, A., 2008. Structured versus unstructured tagging: a case study. Online Inform. Rev. 32 (5), 635–647.

Beißwenger, M., Ermakova, M., Geyken, A., Lemnitzer, L., Storrer, A., 2012. A TEI Schema for the representation of computer-mediated communication. J. Text Encoding Init. (3). Available from: http://jtei.revues.org/476.

Benoit, E.A., 2014. MPLP: A comparison of domain novice and expert user-generated tags in a minimally processed digital archive (Doctoral dissertation). Available from: http://dc.uwm.edu/cgi/viewcontent.cgi?article=1555&context=etd.

Berners-Lee, T., 2006. Linked data. Available from: http://www.w3.org/DesignIssues/LinkedData.html.

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web. Sci. Am. 284 (5), 28–37.

Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked data-the story so far. Int. J. Semant. Web Inf. Syst. 5 (3), 1–22.

Byrne, G., Goddard, L., 2010. The strongest link: libraries and linked data. D-Lib Mag. 16(11/12). Available from: http://www.dlib.org/dlib/november10/byrne/11byrne.print.html.

Cantara, L., 2005. METS: the metadata encoding and transmission standard. CCQ 40 (3–4), 237–253.

CDP Metadata Working Group, 2006. Dublin Core Metadata Best Practices. Version 2.1.1. Available from: http://sustainableheritagenetwork.org/system/files/atoms/file/CDPDublinCoreBPs_0.pdf.

Chan, L.M., Childress, E., Dean, R., O'Neill, E.T., Vizine-Goetz, D., 2001. A faceted approach to subject data in the Dublin Core metadata record. J. Internet Cataloging 4 (1–2), 35–47.

Chopey, M.A., 2005. Planning and implementing a metadata-driven digital repository. CCQ 40 (3/4), 255–287.

Chuttur, M.Y., 2014. Investigating the effect of definitions and best practice guidelines on errors in Dublin Core metadata records. J. Inform. Sci. 40 (1), 28–37.

Clark, J.R., 2008. The Internet connection: Web 2.0, Flickr and endless possibilities. Behav. Soc. Sci. Librar. 27 (1), 62–64.

Coyle, K., 2012. Linked data tools: connecting on the Web. Lib. Technol. Rep. 48, 1–45.

Cummings, J., 2008. The text encoding initiative and the study of literature. A Companion to Digital Literary Studies. Blackwell, Oxford, pp. 451–476.

Cundiff, M.V., 2004. An introduction to the metadata encoding and transmission standard (METS). Lib. Hi Tech. 22 (1), 52–64.

DBpedia., 2015. About. Available from: http://wiki.dbpedia.org/.

DCMI: Dublin Core Metadata Initiative, 2015a. DCMI metadata terms. Available from: http://dublincore.org/documents/dcmi-terms/#H1.

DCMI: Dublin Core Metadata Initiative, 2015b. Metadata basics. Available from: http://dublincore.org/metadata-basics/.

DeRidder, J., Presnell, A., Walker, K., 2012. Leveraging encoded archival description for access to digital content: a cost and usability analysis. Am. Arch. 75 (1), 143–170.

Dillon, M., Jul, E., 1996. Cataloging Internet resources: the convergence of libraries and Internet resources. CCQ 22 (3–4), 197–238.

Ding, Y., Jacob, E.K., Zhang, Z., Foo, S., Yan, E., George, N.L., Guo, L., 2009. Perspectives on social tagging. J. Am. Soc. Inf. Sci. Technol. 60 (12), 2388–2401.

Dulock, M., 2012. Report of the ALCTS metadata interest group meeting, American Library Association Midwinter Meeting, Dallas, January 2012. Tech. Serv. Q. 29(4), 312–317.

Dulock, M., Long, H., 2011. The conference on world affairs archive online: digitization and metadata for a digital audio pilot. D-Lib Mag. 17(3), 3. Available from: http://www.dlib.org/dlib/march11/dulock/03dulock.html.

Duval, E., Hodgins, W., Sutton, S., Weibel, S.L., 2002. Metadata principles and practicalities. D-Lib Mag. 8(4), 16. Available from: http://www.dlib.org/dlib/april02/weibel/04weibel.html.

Elings, M.W., Waibel, G., 2007. Metadata for all: descriptive standards and metadata sharing across libraries, archives and museums. First Monday 12(3). Available from: http://firstmonday.org/article/view/1628/1543.

Eustis, J.M., 2013. Tech services on the Web. Tech. Serv. Q. 30 (4), 441–442.

Flickr: The Commons, 2015. Participating institutions. Available from: https://www.flickr.com/commons.

Foulonneau, M., Riley, J., 2008. Metadata for Digital Resources: Implementation, Systems Design and Interoperability. Chandos Publishing, Oxford.

Getty Research Institute, 2014. Metadata standards crosswalk. Available from: http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html.

Getty Research Institute, 2015. Getty vocabularies as linked open data. Available from: http://www.getty.edu/research/tools/vocabularies/lod/.

Gilliland, A., 2008. Setting the stage. In: Baca, M. (Ed.), Introduction to Metadata (1–19). The Getty Research Institute, Los Angeles (CA).

Gilliland-Swetland, A.J., 1998. An exploration of K-12 user needs for digital primary source materials. Am. Arch. 61 (1), 136–157.

Guenther, R.S., 2003. MODS: the metadata object description schema. Portal Lib. Acad. 3 (1), 137–150.

Guenther, R.S., 2004. Using the metadata object description schema (MODS) for resource description: guidelines and applications. Lib. Hi Tech. 22 (1), 89–98.

Guy, M., Tonkin, E., 2006. Tidying up tags. D-Lib Mag. 12 (1). Available from: http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january06/guy/01guy.html.

Hammond, T., Hannay, T., Lund, B., Scott, J., 2005. Social bookmarking tools (I) a general review. D-Lib Mag. 2 (4). Available from: http://www.dlib.org/dlib/april05/hammond/04hammond.html.

Hatop, G., 2013. Integrating linked data into discovery. Code4Lib J. 21, 67–90. Available from: http://journal.code4lib.org/articles/8526?utm_source=rss&utm_medium=rss&utm_campaign=integrating-linked-data-into-discovery.

Hedden, H., 2008. Controlled vocabularies, thesauri, and taxonomies. Indexer 26 (1), 33–34.

Hillmann, D.I., 2008. Metadata quality: from evaluation to augmentation. CCQ 46 (1), 65–80.

Huang, H., Jörgensen, C., 2013. Characterizing user tagging and co-occurring metadata in general and specialized metadata collections. J. Am. Soc. Inf. Sci. Technol. 64 (9), 1878–1889.

Jeong, W., 2009. Is tagging effective?–overlapping ratios with other metadata fields. In: International Conference on Dublin Core and Metadata Applications, Seoul, Korea, 12–16 October, pp. 31–39.

Kennedy, M.R., 2008. Nine questions to guide you in choosing a metadata schema. J. Digital Inf. 9 (1).

Kipp, M.E., 2011. Controlled vocabularies and tags: an analysis of research methods. NASKO 3 (1), 23–32.

Kucsma, J., Reiss, K., Sidman, A., 2010. Using Omeka to build digital collections: the METRO case study. D-Lib Mag. 16 (3/4). Available from: http://www.dlib.org/dlib/march10/kucsma/03kucsma.html.

Kurtz, M., 2013. Dublin Core, DSpace, and a brief analysis of three university repositories. Inf. Technol. Lib. 29 (1), 40–46.

Lagoze, C., Lynch, C.A., Daniel Jr, R., 1996. The Warwick Framework: a container architecture for aggregating sets of metadata. Cornell University, Ithaca, NY. Available from: https://ecommons.cornell.edu/handle/1813/7248.

Lampert, C.K., Southwick, S.B., 2013. Leading to linking: introducing linked data to academic library digital collections. J. Lib. Metadata 13 (2–3), 230–253.

Lange, H.R., Winkler, B.J., 1997. Taming the Internet. Advances in Librarianship, Vol. 21. Emerald Group Publishing Limited, Bingley, UK, 21, 47–72.

Laursen, D., Christiansen, K.F., Olsen, L.L., 2012. Management of metadata for digital heritage collections. Microform Digit. Rev. 41 (3/4), 151–158.

Leise, F., 2008. Controlled vocabularies, an introduction. Indexer 26 (3), 121–126.

Library of Congress, 2011a. MIX (NISO Metadata for Images in XML). Available from: http://www.loc.gov/standards/mix//.

Library of Congress, 2011b. AudioMD and VideoMD—Technical metadata for audio and video. Available from: http://www.loc.gov/standards/amdvmd/index.html.

Library of Congress, 2013. EAD: Encoded Archival Description. Version 2002. Available from: http://www.loc.gov/ead/.

Library of Congress, 2014a. MODS User Guidelines version 3. Available from: http://www.loc.gov/standards/mods/userguide/.

Library of Congress, 2014b. VRA Core. Available from: http://www.loc.gov/standards/vracore/.

Library of Congress, 2015a. Metadata Object Description Schema (MODS). Available from: http://www.loc.gov/standards/mods/.

Library of Congress, 2015b. METS: Metadata Encoding and Transmission Standard. Available from: http://www.loc.gov/standards/mets/.

Library of Congress, 2015c. Metadata Object Description Schema (MODS). Conversion. Available from: http://www.loc.gov/standards/mods/mods-conversions.html.

Library of Congress, 2015d. LC linked data service: authorities and vocabularies. Available from: http://id.loc.gov/.

LODLAM: Linked open data in libraries, archives, and museums, 2015. About. Available from: http://lodlam.net/about/.

Lopatin, L., 2010. Metadata practices in academic and non-academic libraries for digital projects: a survey. CCQ 48 (8), 716–742.

Lu, C., Park, J.R., Hu, X., 2010. User tags versus expert-assigned subject terms: a comparison of LibraryThing tags and Library of Congress Subject Headings. J. Inf. Sci. 36 (6), 763–779.

Macgregor, G., McCulloch, E., 2006. Collaborative tagging as a knowledge organisation and resource discovery tool. Lib. Rev. 55 (5), 291–300.

Matusiak, K.K., 2006. Towards user-centered indexing in digital image collections. OCLC Syst. Serv. Int. Digital Lib. Persp. 22 (4), 283–298.

Matusiak, K.K., Johnston, T., 2014. Digitization for preservation and access: restoring the usefulness of the nitrate negative collections at the American Geographical Society Library. Am. Arch. 77 (1), 241–269.

McCallum, S.H., 2004. An introduction to the Metadata Object Description Schema (MODS). Lib. Hi Tech. 22 (1), 82–88.

McCrory, A., Russell, B.M., 2013. Crosswalking EAD: collaboration in archival description. Inf. Technol. Lib. 24 (3), 99–106.

McDonough, J.P., 2006. METS: standardized encoding for digital library objects. Int. J. Digital Lib. 6 (2), 148–158.

Merholz, P., 2004. Metadata for the masses. Available from: http://www.adaptivepath.com/ideas/e000361/.

Miller, S.J., 2011. Metadata for Digital Collections: A how-to-do-it Manual. Neal-Schuman Publishers, New York, N.Y.

Miller, S.J., 2015. Ontologies for semantic applications. In: Smiraglia, R., Lee, H.L. (Eds.), Ontology for Knowledge Organization. Ergon Verlag, Wurzburg, Germany, pp. 87–106.

Mitchell, E.T., 2013a. Metadata developments in libraries and other cultural heritage institutions. Lib. Technol. Rep. 49 (5), 5–10.

Mitchell, E.T., 2013b. Building blocks of linked open data in libraries. Lib. Technol. Rep. 49 (5), 11–25.

Mitchell, E.T., 2013c. Three case studies in linked open data. Lib. Technol. Rep. 49 (5), 26–43.

Mixter, J., 2014. Using a common model: mapping VRA Core 4.0 into an RDF ontology. J. Lib. Metadata 14 (1), 1–23.

Moulaison, H.L., Dykas, F., Gallant, K., 2015. OpenDOAR repositories and metadata practices. D-Lib Mag., 21 (3/4), 1. Available from: http://www.dlib.org/dlib/march15/moulaison/03moulaison.html.

NISO, Framework Working Group, 2007. A Framework of Guidance for Building Good Digital Collections, 3rd ed. Available from: http://www.niso.org/publications/rp/framework3.pdf.

NISO National Information Standards Organization, 2004. Understanding Metadata. http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

Niu, J., 2013. Hierarchical relationships in the bibliographic universe. CCQ 51 (5), 473–490.

OAI: Open Archives Initiative, 2015. Standards for web content interoperability. Available from: https://www.openarchives.org/.

Ore, C.E., Eide, Ø., 2009. TEI and cultural heritage ontologies: exchange of information? Lit. Ling. Comp. 24 (2), 161–172.

Palmer, C.L., Zavalina, O.L., Mustafoff, M., 2007. Trends in metadata practices: a longitudinal study of collection federation. In: Proceedings of the Seventh ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 386–395.

Park, J.R., 2009. Metadata quality in digital repositories: a survey of the current state of the art. CCQ 47 (3–4), 213–228.

Park, J.-R., Tosaka's, Y., 2010. Metadata creation practices in digital repositories and collections: schemata, selection criteria, and interoperability. Inf. Technol. Lib. 29 (3), 104–116.

Pattuelli, C., Rubinow, S., 2013. The knowledge organization of DBpedia: a case study. J. Document. 69 (6), 762–772.

Pattuelli, M.C., Miller, M., Lange, L., Fitzell, S., Li-Madeo, C., 2013. Crafting linked open data for cultural heritage: mapping and curation tools for the linked jazz project. Code4Lib J. 21.

Petek, M., 2012. Comparing user-generated and librarian-generated metadata on digital images. OCLC Syst. Serv. Int. Digital Lib. Persp. 28 (2), 101–111.

Pirmann, C., 2012. Tags in the catalogue: insights from a usability study of LibraryThing for libraries. Lib. Trends 61 (1), 234–247.

Ransom, N., Rafferty, P., 2011. Facets of user-assigned tags and their effectiveness in image retrieval. J. Document. 67 (6), 1038–1066.

Rorissa, A., 2010. A comparative study of Flickr tags and index terms in a general image collection. J. Am. Soc. Inf. Sci. Technol. 61 (11), 2230–2242.

Singer, R., 2009. Linked library data now! J. Electron. Res. Librar. 21 (2), 114–126.

Southwick, S.B., 2015. A guide for transforming digital collections metadata into linked data using open source technologies. J. Lib. Metadata 15 (1), 1–35.

Sperberg-McQueen, C.M., 1996. Textual criticism and the text encoding initiative. In: Finneran, R.J. (Ed.), The Literary Text in the Digital Age. University of Michigan Press, Ann Arbor, MI, pp. 37–61.

Sperberg-McQueen, C.M., Burnard, L., 1994. A gentle introduction to SGML. In: Guidelines for Electronic Text Encoding and Interchange. TEI Working Committees. University of Michigan Libraries, Chapter 2. Available from: http://quod.lib.umich.edu/t/tei/

Taylor, A.G., Joudrey, D.N., 2008. The Organization of Information. Westport, CT: Libraries Unlimited.

TEI Consortium, 2015. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Available from: http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf.

Thomas, M., Caudle, D.M., Schmitz, C., 2010. Trashy tags: problematic tags in LibraryThing. New Lib. World 111 (5/6), 223–235.

Thomas, M., Caudle, D.M., Schmitz, C.M., 2009. To tag or not to tag? Lib. Hi Tech. 27 (3), 411–434.

Trant, J., 2009. Tagging, folksonomy and art museums: early experiments and ongoing research. J. Digital Inf. 10 (1.).

Van Assem, M., Van Ossenbruggen, J., Schreiber, G., 2010. The VRA core application profile for searching and presenting cultural heritage: the MultimediaN case. In: Proceedings of the International Conference on Dublin Core and Metadata Applications, Pittsburgh, PA, USA.

Van Hooland, S., Verborgh, R., 2014. Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata. Neal-Schuman, Chicago.

Vanhoutte, E., 2004. An introduction to the TEI and the TEI Consortium. Lit. Ling. Comp. 19 (1), 9–16.

Vellucci, S.L., 1998. Metadata. Annual Review of Information Science and Technology (ARIST), 33, 187–222.

Visual Resources Association, 2007. VRA Core 4.0 element description. Available from: https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf.

W3C RDF Working Group, 2014. RDF 1.1 Primer. Available from: http://www.w3.org/TR/rdf11-primer/.

W3C: World Wide Web Consortium, 2011. Library Linked Data Incubator Group Final Report. Available from: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/.

W3C: World Wide Web Consortium, 2015a. Semantic Web. Available from: http://www.w3.org/standards/semanticweb/.

W3C: World Wide Web Consortium, 2015b. Linked Data. Available from: http://www.w3.org/standards/semanticweb/data.

Weibel, S.L., Koch, T., 2000. The Dublin core metadata initiative. D-Lib Mag. 6 (12), Available from: http://mirror.dlib.org/dlib/december00/weibel/12weibel.html.

Wisneski, R., Dressler, V., 2009. Implementing TEI projects and accompanying metadata for small libraries: rationale and best practices. J. Lib. Metadata 9 (3–4), 264–288.

Woodley, M.S., 2008. Crosswalks, metadata harvesting, federated searching, metasearching: using metadata to connect uses and information. In: Baca, M. (Ed.), Introduction to Metadata. The Getty Research Institute, Los Angeles (CA), pp. 38–62.

Yaco, S., 2008. It's complicated: barriers to EAD implementation. Am. Arch. 71 (2), 456–475.

Yakel, E., 2004. Encoded archival description: are finding aids boundary spanners or barriers for users? J. Arch. Org. 2 (1–2), 63–77.

Yakel, E., Kim, J., 2005. Adoption and diffusion of Encoded Archival Description. J. Am. Soc. Inf. Sci. Technol. 56 (13), 1427–1437.

Yoose, B., Perkins, J., 2013. The linked open data landscape in libraries and beyond. J. Lib. Metadata 13 (2–3), 197–211.

Zeng, M.L., Qin, J., 2008. Metadata. Neal-Schuman Publishers, New York, NY.