Metadata

Marcia Lei Zeng

and

Jian Qin

Neal-Schuman Publishers, Inc.
New York London

Published by Neal-Schuman Publishers, Inc.
100 William St., Suite 2004
New York, NY 10038

Printed and bound in the United States of America

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

Library of Congress Cataloging-in-Publication Data

Zeng, Marcia Lei, 1956–
Metadata / Marcia Lei Zeng and Jian Qin.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-55570-635-7 (alk. paper)
1. Metadata. I. Qin, Jian, 1956– II. Title.

Z666.7Z46 2008
025.3—dc22

2008015176

Don’t miss the companion Web site that accompanies this book available at:

www.neal-schuman.com/metadata/

Contents

Part II. Metadata Building Blocks

Part III. Metadata Services

List of Illustrations

Figures

Exhibits

Tables

Preface

Although rooted in library and information science (the first metadata scheme targeted for Internet resources—the Dublin Core Metadata Element Set—was proposed in 1995), metadata has expanded its territory beyond traditional libraries and is now a widely adopted vital solution for describing the explosively growing, complex world of digital information.

As many organizations turn to metadata applications for managing massive quantities of digital information, demand increases for information professionals who are prepared for the immediate tasks at hand. During the past decade, this book's authors have engaged in teaching metadata and information organization courses as well as conducting research in this area. The authors have also had opportunities to provide training for professionals and to act as consultants for digital library projects. The experience we accumulated through teaching, research, and consulting motivated us to write a textbook that systematically introduces metadata concepts and principles through the incorporation of practical examples and learning assessment materials.

Metadata is both a textbook and an instructional guide for practitioners. As a textbook, its primary purpose is to provide educators with a convenient and reliable source for teaching metadata-related courses in universities or in continuing education programs for information professionals. Among the unique features of this book are instructional materials such as sample problems and solutions and hands-on tutorials. These instructional features also make *Metadata* an ideal resource for practitioners who wish to use the book for self-study or on-the-job training.

While focusing on metadata concepts, principles, and applications, the book also covers trends, innovative ideas, and advanced technologies in metadata research and practice that will have significant implications in the years to come. The wide application of metadata in different domains has created different communities of practice, each of which defines a metadata structure based on its own norms and needs. We will therefore cover the conceptual and practical knowledge that is fundamental to all application domains. This is not an overview of all existing metadata standards, nor is it an interpretation of individual metadata schemas. Although many of those will be mentioned or discussed and their features will be referenced as examples, the text is not a step-by-step manual for creating metadata records. Rather, it identifies commonalities among metadata schemas and focuses on the design and profiling processes as they relate to the needs of application domains and environments. The inclusion or exclusion of a schema in this book should not be interpreted as a sign of favoritism or preference for one schema over another.

Focus and Organization

The topics covered in this book are selected and organized based on an outcome-oriented learning philosophy which holds that regardless of learners' locations or backgrounds, we can expect them to be able to learn the how-tos of metadata application design, implementation, and evaluation, in addition to understanding the underpinning theory. This approach allows learners of all kinds and skill levels to adapt the knowledge and practices they obtain from this book to the domains in which they work. Therefore, we concentrate on the tasks typical to successful implementation of metadata application projects. Such tasks include applying an existing standard to a project, establishing localized element sets or application profiles by drawing elements from multiple metadata schemes, and performing advanced tasks related to services, integration, and assessment.

Metadata is divided into four parts: "Fundamentals of Metadata," "Metadata Building Blocks," "Metadata Services," and "Metadata Outlook in Research."

The first part includes Chapters 1 and 2. Starting with metadata application scenarios, Chapter 1 introduces the definitions, types and functions, principles, and anatomy of metadata. It provides a bridge for readers from abstract scenarios to real-world applications of metadata functions and structures in digital environments. Chapter 2 introduces metadata standards within major application domains. We emphasize semantics of the element sets, the needs of domain-specific information objects, and the functions these standards aim to fulfill. Standards covered in Chapter 2 include those for general purposes, cultural objects and visual resources, learning objects, archives and preservation, rights management, scientific data, media, and agents.

The second part of the book moves from general fundamentals to metadata building blocks. Chapter 3 is devoted to the development of the structure and semantics of a metadata schema. It discusses perspectives and techniques for assessing needs in different project environments: identifying desired elements and refinements for an element set, controlling the values in value spaces, creating application profiles, and establishing crosswalks between or among element sets. The last section explains what should be included in best practice guidelines and how guidelines should be presented.

While Chapter 3 discusses how semantics in metadata elements and their structures are defined, Chapter 4 details how the schema will be encoded and how the semantics are controlled by using namespaces. XML Schemas from a flat structure element set, a hierarchical structure set, and an application profile provide useful examples. The goal and outcome is to provide a basic understanding of the issues that may arise in applications regarding schema encoding.

Chapter 5 is a major component of the text, dedicated to the issues and techniques related to creating metadata records. It can be considered as consisting of two major topics: the issues related to metadata records and the issues regarding encoding. Conceptual models are first presented to provide a better understanding about metadata statements that form the descriptions of the resources. The discussion then turns to the granularity of records, i.e., levels of description at which a metadata record may be created. We emphasize creating sharable records because interoperability is an important concept in metadata applications. Discussion of metadata resources and tools presents the options that records may be created by human catalogers fully or partially, generated by computer programs, or converted and harvested from other sources. Chapter 5 gives these issues a closer examination. Encoding metadata is a long and comprehensive section, in which metadata storage methods are introduced first, followed by details and examples of expressing metadata records in HTML/XHTML, XML, and RDF. The last section covers other methods related to metadata records, such as linkage, wrapper, display, and parallel metadata.

The third part of the book brings together metadata services that have appeared in recent years. Chapter 6 introduces the types of infrastructures for these services. Standards in XML, RDF, data communication, policies, and procedures promise an exciting yet challenging future for metadata services. Detailed explanations are applied to metadata registries and repositories (including the metadata harvesting protocol). For each of these services, we look at the functionalities of the service, basic components, and types of models so that learners gain a basic understanding of these advanced topics. Chapter 6 summarizes the emerging approaches to ensure optimal metadata discovery through discussions involving metadata retrieval technologies and methods of exposing metadata and maximizing its usage. Chapter 7 offers a systematic view of the issues and methodologies of measuring metadata quality. Evaluation criteria, measurement processes, and methods of evaluation are discussed in detail. Chapter 8 summarizes the methods of ensuring and achieving interoperability based on research of this all-important issue. Interoperability approaches are analyzed at the schema, record, and service levels. Examples are selected from projects throughout the world.

The final section draws attention to the research landscape. Chapter 9 reviews major research areas in metadata architecture, modeling, and metadata semantics that are not discussed in detail in the rest of the text.

As professional educators, we understand the importance of exercises and practical examples in a textbook. Each chapter in this book provides a recommended reading list, some with a series of practice and assessment instruments. In addition to general exercises at the end of major chapters, the digital library prepared for instructors contains detailed exercises and hints for some assignments. All exercises have been created as an interactive component, available either on an instructor's CD-ROM from the publisher or via this book's accompanying Web site. We hope that our experience in metadata research, teaching, and consulting will offer our readers a unique, enlightening, and holistic approach to the topic.

Part I
Fundamentals of Metadata

1

Introduction

Digital information exists in every area of modern life: in the office, at home, and on the road. Every organization and individual faces the challenge of organizing digital information of all types and formats, while also being able to find the correct or desired information in a timely manner (and with acceptable precision of sufficient depth). For a long time, libraries and information database producers were the primary agents organizing and providing both information and efficient search tools. The techniques and technologies they used are often proprietary and idiosyncratic. Along with the rise of the Internet, Web-based technologies enabled the creation of mass information and publication through a low-barrier platform—anyone who can use a text or image editor can now create digital documents or objects and publish them directly on the Web. Such technology greatly democratized the publication and dissemination of information and resulted in an exponentially faster increase in the volume and complexities of digital resources. Individuals, organizations, communities, and governments now face the tasks of organizing the massive amounts of digital information in their systems before they can effectively discover, locate, and use it when needed.

At present, the Internet and the World Wide Web (WWW) have become the new library catalogs, indexing databases, dictionaries, encyclopedias, newspapers, schools, entertainment centers, and many other sources and places that we used to *physically* access. How do we find the sources we need and the places we want to go on the Web or within an internal Web site? We use search engines, of course. However, how do they take us where we need to go in the world of digital resources? What makes search engines work? Moreover, what makes them work effectively? The "invisible hand" of efficient organization is embodied in *metadata*. When one types a keyword into a major search engine or, better yet, into a digital library's search field, the chances are good that the keyword is one of thousands in an index. The index may be compiled from a metadata repository or extracted from full-text documents. The document information displayed in the search results pages is a typical instantiation of metadata: it describes what a document is, what it is about, and where one may locate it. Let us examine the following two examples of metadata application.

A Maths Dictionary for Kids (www.amathsdictionaryforkids.com/maths/dictionary.html, accessed 2007) is an interactive tool designed for K–6 grade-schoolers to learn mathematical concepts by allowing children to follow the instructions and practice math problem solving. Teachers preparing lessons may incorporate the interactive exercise for math games into instructional materials; they can then search learning objects by topic, e.g., "math," or by form, e.g., "interactive games," and combine different criteria to filter out irrelevant information. In this case, *grade level*, *subject*, *form*, and other metadata fields enhance the precision of a query in a system already designed to produce more relevant results.

Today, a search in a digital library produces hundreds of displayed results. The user then must wade through a morass in order to select relevant information and may eventually give up in frustration. Here, the "tolerance for ambiguity" (the patience, perseverance, and equanimity necessary when an answer, a resource, or the results of one's work are not immediately forthcoming)—which is the hallmark of the researcher—is not commonly found with everyday users. The use of metadata, however, allows the system to perform post-search processing and present the results in categorized groups. As Figure 1-0-1 shows, the search system in the Gateway to Educational Materials (GEM: www.thegateway.org) provides a categorized list of clustered results (at the lower right-hand corner) by using the subject terms assigned to metadata records. Among the 85 items matching the search query *math lesson plan*, four are in *arts*, two are in *educational technology*, 24 are in *mathematics*, and so forth.

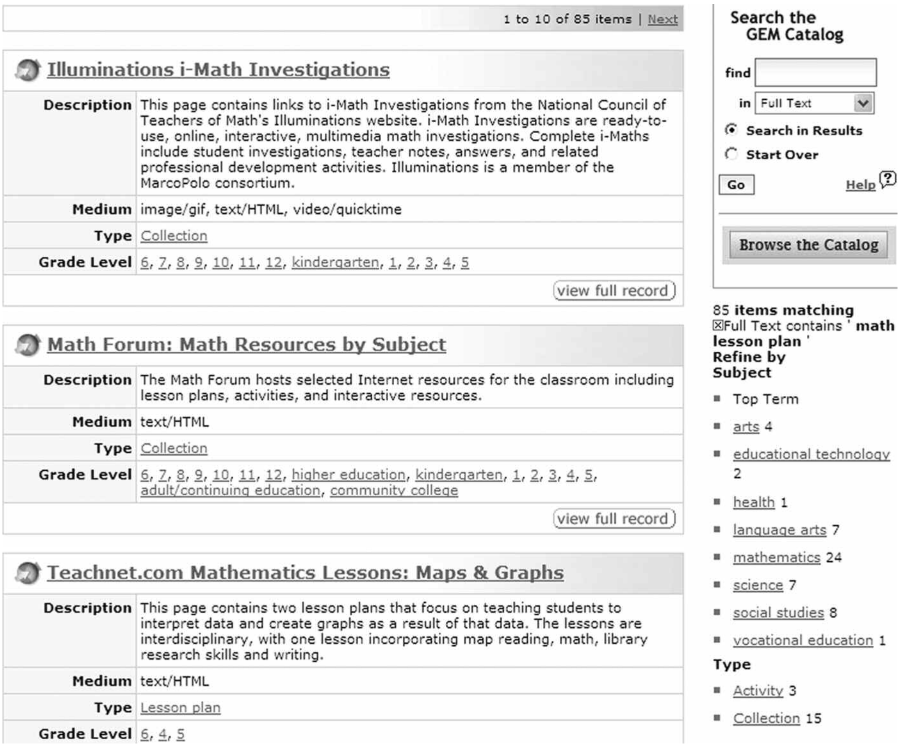
The previous examples demonstrate that metadata is capable of performing the following tasks:

- Describing what resources are and what they are about, and organizing those resources according to controllable criteria.
- Allowing resources to be found by relevant criteria, aggregating similar resources, and providing pathways to the location of the desired information.
- Facilitating metadata exchange and enabling interoperability.
- Providing digital identification and description for archiving and the preservation of resources (NISO, 2004).

1.1 A Brief History

The organization of objects and phenomena into either classes or sets of relationships is one way in which humans communicate. Before the Internet, organizing information in libraries, archives, museums, and other types of institutions was governed by highly structured rules and standards such as the *Anglo-American Cataloguing Rules Revised*, Second Edition (AACR2) and MARC (*MAchine-Readable Cataloging*). The content representation was guided by semantically rich classification schemes and lists of subject headings, e.g., the *Dewey Decimal Classification* (DDC) and the *Library of Congress Subject Headings* (LCSH). The information objects being organized were primarily physical, i.e., information was packaged in some sort of container or packaged as books,

Figure 1-0-1. Categorized search result display in GEM.



Source: Courtesy of the GEM Exchange Managers for the Gateway to 21st Century Skills, www.thegateway.org, accessed 2007.

journals, CD-ROMs, audio/video cassettes, film reels, and more. A large amount of human catalog work and intervention was involved in the process, not only because of the physical nature of information objects but also because complex rules and standards governed the relationships among these objects. All of these factors made it very difficult, if not impossible, for computer programs to take over metadata creation tasks from human catalogers. However, the level of sophistication and maturity of information technology also played an important role in pre-WWW cataloging as evidenced by the production of MARC-compliant AACR catalog cards then mass-produced in Dublin, Ohio, by the Online Library Computer Center (OCLC). OCLC remains the central authority for all descriptive and technical cataloging that provides bibliographic and other formatted packages that Online Public Access Catalogs (OPACs) allow users to access in all libraries.

Pre-Internet cataloging (e.g., typewritten cards) played a significant role in assisting users to find what they needed *and* to know whether an item was located in the stacks, and whether collocated items of the same subject area made the adventure into the musty rows of books worthwhile. The purposes of pre-Internet cataloging were twofold: (1) to provide rich bibliographic descriptions and relationships between and among data of heterogeneous items and (2) to facilitate sharing these bibliographic data across library boundaries. While AACR2 and MARC have done a meritorious job in accomplishing those purposes, they have fallen short on several important fronts in Internet-based resource descriptions, e.g., management of digital rights, preservation of digital objects, and evaluation of resources based on authenticity, user profile, and grade level.

Metadata development in the Internet era took off in the first half of the 1990s when the Internet was becoming a "household" word for libraries and institutions that manage and use large amounts of digital information on a daily basis. During that period, distributed information repositories over the Internet grew at an exponential rate. The chaotic Internet-based information called for mechanisms of description, authentication, and management, which prompted the development of new guidelines and architectures by different communities. Several parallel development areas in metadata existed in the early 1990s. The scientific community started looking for solutions to organize the rapidly increasing scientific data, which prompted the debut of the *Content Standards for Digital Geospatial Metadata* (CSDGM) in 1992 by the Federal Geographic Data Committee (FGDC). In the humanities community, the Text Encoding Initiative (TEI), an international organization founded in 1987, released the first version of *Guidelines for Electronic Text Encoding and Interchange* (TEI Guidelines) in 1994. As an international and interdisciplinary standard, TEI Guidelines "focus (though not exclusively) on the encoding of documents in the humanities and social sciences, and in particular on the representation of primary source materials for research and analysis" (TEI: www.tei-c.org/Guidelines/index.xml, accessed 2007).

The library community also took action to develop metadata standards as a solution to resource description and discovery problems. OCLC initiated a project in 1994 to experiment with cataloging Web resources by using the AACR2 and MARC format. Over 200 volunteer librarians created more than 2,500 records for Internet resources (1995), which became the precursor of the invitational Metadata Workshop held in 1995 at OCLC in Dublin, Ohio. The *Dublin Core* was born and emerged from this historically significant workshop. Following the first Dublin Core workshop, a metadata movement soon spread rapidly to other continents and across research, educational, and governmental institutions, as well as businesses, and many organizations. The following represent some of the many metadata standards developed since the 1990s:

- Categories for the Description of Works of Art (CDWA)
- Visual Resources Association (VRA) Core Categories
- Learning Object Metadata (LOM), Institute of Electrical and Electronics Engineers (IEEE)
- Encoded Archival Description (EAD)
- Metadata Object Description Schema (MODS)

- PREMIS: PREservation Metadata Implementation Strategies
- ONline Information Exchange (ONIX)
- Digital Object Identifier (DOI)
- The Friend of a Friend (FOAF)
- MPEG-7 (the standard for description and search of audio and visual content)
- Public Broadcasting Metadata Dictionary (PBCore)

Metadata projects increased greatly during the late 1990s. The DCMI (Dublin Core Metadata Initiative) home page formerly maintained a list of metadata projects and the number quickly increased into the hundreds, gathered from all across the world. Publications about metadata dominated Internet forums, professional journals, and conferences. The main reason for this proliferation is because no limit exists for the type or amount of resources that metadata can describe, nor are there any limits to the number of overlapping standards for any type of resource or subject domain. The past decade witnessed a continual expansion and evolution of metadata research and practices at almost all levels and areas where human activities and digital information converge.

1.2 Definitions

The term *metadata* was used before the advent of the Internet, primarily in computer science to specify information about database objects and/or program objects. Broadly speaking, metadata encapsulates the information that describes any document or object in both digital and traditional formats. Metadata is often simply defined as "data about data" or "information about information." As research and applications have evolved, metadata has been refined to "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (NISO, 2004: 1). The DCMI defines metadata as "data associated with either an information system or an information object for purposes of description, administration, legal requirements, technical functionality, use and usage, and preservation" (DCMI Glossary, 2005). *Metadata*, like *data*, is expressed as singular or plural. It is usually singular in the sense of a kind of data, but when referring to values in metadata statements, the term designates things one can count (Turner, 2003; NJ-BGIS, 2007). It is also referred to in the literature as "meta data" and "meta-data." In this text we use more precise phrases for the many related concepts that are sometimes mistakenly used, typically a *metadata statement* (e.g., <title = "ABC">) or a *metadata record* (i.e., a particular instance that a set of metadata elements is applied to for describing an object) is abbreviated as "metadata." Several of these concepts are discussed in Chapter 5, together with the abstract models that have been developed by the metadata community.

A great number of metadata standards have been developed or proposed by different communities and subject domains. A key component in these standards is the *element set* (more appropriately, the *scheme*) that defines the structure and semantics of elements. For example, the international standard *Dublin*

Core Metadata Element Set (DCMES) defines 15 core elements that should be used to describe distributed information resources on the Internet for discovery purposes. The metadata community has used a number of terms in different contexts to refer to metadata *element sets*, in addition to the ways that they are represented in machine-processable formats. The terms "metadata schemes," "metadata standards," "element sets," and "metadata schemas" have subtle differences but are used interchangeably in the literature. These phrases are explained and differentiated in the text where necessary.

Components of metadata have also been addressed differently by various communities. For example, a database designer may call metadata elements "data fields" instead of "elements," while a metadata developer may see metadata as an abstract model that can be implemented through a schema and a set of guidelines. In this case, the abstract model would explain the concepts and relationships between classes of concepts and properties in the domain for which the schema is to be designed, whereas the schema itself would define a set of elements, attributes, and rules for value spaces.

1.3 Types and Functions

Metadata applications use metadata records to describe a resource by recording *title*, *creator*, *keywords*, and other information, which is analogous to the technical cataloging of a library. However, several important characteristics make it distinctive from the traditional catalog. First, the administrative function of metadata is facilitated by knowing when and how a record was created (by whom and from where), what technical details it contains, and who has access privileges. Rights management and preservation are the most important types of *administrative metadata*. Digital resources can often be easily accessed, copied, modified, or deleted, which in turn can trigger violations of copyright, access permissions, and licensing rules. In addition, serious difficulties may develop in preserving digital resources for use by future generations of software. Administrative metadata will record information on rights and the technical characteristics of a resource, which will be used for organized administrative tasks and management. In addition to describing the resource's characteristics, the metadata may also be used to describe how this resource should be or has been used, e.g., what platform and software is required to run a digital object, and what are the targeted ages or grade levels of a learning object.

Metadata is different from library cataloging because resources that it works with extend beyond traditional format. Contemporary metadata, increasingly concerned with digital resources, brings two major benefits: (1) users can navigate directly from a metadata record (i.e., a surrogate of a resource) to the resource on the Internet or point to where its physical location might be; and (2) "automatic processability," which allows a metadata creation agency to process certain types of resources directly through computer programs so that the metadata records (or a portion of those records) may be generated by automatic (computer program) processes. Automatic metadata generation indicates a minimal amount of (or none) human intervention, and so is cost-efficient for metadata records creation. Library materials, on the other hand, are dependent upon human catalogers to record the data manually into a rigid format, typically the MARC format.

Defining types of metadata is both contextual and dependent upon application domains. The Getty Information Institute (1983–1999), now the Getty Research Institute, of the J. Paul Getty Trust published a well-known book *Introduction to Metadata: Pathways to Digital Information* in 1998 (Gill et al., 1998) and made its online versions (Baca, 2000–2008). It identifies five types of metadata and their functions: *administrative*, *descriptive*, *preservation*, *technical*, and *use* metadata (Gilliland, 2000). In a booklet published by the National Information Standards Organization (NISO), metadata types are given three summative groupings that have been widely accepted by the metadata community:

- Descriptive metadata describes a resource for purposes such as discovery and identification;
- Structural metadata indicates how compound objects are put together;
- Administrative metadata provides information to help manage a resource, and includes technical, rights management, and preservation metadata. (NISO, 2004: 1)

Organizing digital resources and providing services for retrieving and using them is a complex process that requires various metadata used for different purposes and functions. The most essential display and describe the digital resource. The metadata community typically calls this the *description* of resources. The validity of classic data elements used in cataloging and indexing will be found here: *title*, *author*, *abstract*, *keywords*, *publisher*, and *date*.

If a digital resource consists of multiple objects or parts, the data elements should be able to capture such information, especially when the sequence and levels of the objects are the most important. A typical example is the digitization of manuscripts, in which each page is often scanned as an image. The *structural* metadata is critical for keeping the content as represented in scanned images in the correct sequence. *Technical* metadata describes when and how the digitized manuscript was created, as well as its file type and other related information. *Rights management* metadata is concerned with intellectual property rights of the manuscript both in its original form and in the digitized form, and *preservation* metadata contains information needed to archive and preserve the manuscript images. All of these types of metadata will offer functions necessary for resource discovery, organizing electronic resources, facilitating interoperability, digital identification, and archiving and preservation (NISO, 2004).

1.4 Principles

The Dublin Core Metadata Initiative was motivated by the guiding principle of producing a metadata element set simple enough for creating and maintaining metadata records, and thus the elements would also conform to existing and emerging standards at the international scope, and also be interoperable among collections and indexing systems (Weibel and Hakala, 1998). These requirements translate into three primary principles for the construction of ideal metadata: *simplicity*, *extensibility*, and *interoperability*. By simply taking only those data elements necessary and so maintaining a minimum set of elements for easy deployment, it must be equally important for the metadata to be flexible enough to accommodate specialized needs. This suggests that metadata schemas should allow applications to introduce new elements and constraints for localized description needs. Extensibility is generally understood in two ways: (1) the ability of a metadata schema to offer a core set of elements that will unify different models of resource description; and (2) the ability to link a simple metadata record to a richer, more complex description of resources (Dempsey and Weibel, 1996). Extensibility requires metadata systems to allow for the addition of new elements and/or subelements to existing ones in a schema, while the new elements are selected from existing metadata standards or can be established at the local level. Interoperability is defined as "the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality" (NISO, 2004: 2).

As metadata development evolved, the early requirements for metadata were eventually extended and elucidated to be a more inclusive and refined set of principles. Duval et al. (2002) pointed out, "*Principles* are those concepts judged to be common to all domains of metadata and which might inform the design of any metadata schema or application. *Practicalities* are the rules of thumb, constraints, and infrastructure issues that emerge from bringing theory into practice in the form of useful and sustainable systems" (Duval et al., 2002: www.dlib.org/dlib/april02/weibel/04weibel.html, accessed 2007). In their paper, metadata principles were extended to also include modularity, refinement, and multilingualism. *Modularity* refers to building metadata into *blocks*, so that data elements, vocabularies, and other building blocks in different metadata schemas may be assembled into new schemas in a syntactically and semantically interoperable way. Different metadata modules, e.g., for discovery, rights management, preservation, or instructional management, expressed in a common syntactic idiom such as Extensible Markup Language (XML), should be able to be combined in compound schemas as needed, which would embody the functionality of each constituent. *Refinement* aims at precise detail that determines how much description a schema should require. *Multilingualism* focuses on aspects of language and culture, and when adopting metadata architecture, the designer needs to take into account linguistic and cultural diversity (Duval et al., 2002). These principles address the issues one may encounter in one form or another in both metadata schema design and record generation processes. As principles, they represent a set of basics that all metadata designers should consider when creating each project. These principles also have a direct effect on how to implement metadata projects and simultaneously make them sustainable for long-term utilization, preservation, and interoperable for sharing and reuse.

1.5 Anatomy of a Metadata Standard

Metadata standards provide guidelines regarding structure, values, and content. These are also the basis for developing software programs and tools that can lead to good descriptive cataloging, consistent documentation, shared records, and increased end-user access (Baca et al., 2006). Section 1.1 provided a list of well-known metadata standards. What are these components and what roles do they play in formulating a schema and structuring a metadata record? Let us take a closer look at the inside of a metadata standard.

The backbone of a metadata standard consists of a set of elements. Each element is given an explanatory name, label, and definition. Sometimes a note (annotation) is provided to document the change history of the element and the mapping between this and similar elements in other standards. The Dublin Core Metadata Element Set provides *name*, *URI* (Uniform Resource Indicator), *label*, *definition*, *comment*, and *references*, for example, for the element *date* (see Figure 1-5-1).

Elements in a metadata standard may be structured in different ways and presented in an *element set* scheme. For example, a standard may simply put all elements into a linear list, i.e., a flat structure, while others may use a hierarchical structure to indicate a parent-child relationship between elements. Labeling elements often follows some naming conventions, i.e., whether elements need to be capitalized, how compound-word elements should be used, and when abbreviations are appropriate.

The next component of a standard is a set of constraints and guidelines for how data values should be recorded when generating metadata records. The standard specifications usually provide, in addition to element definitions, format rules for the types of values (e.g., number, text, date, and so on) that should be consistently applied to all elements. Taking the *date* element defined in DCMES (see Figure 1-5-1) requires an indication of how a date, e.g., date=”050603”, should be formulated as “yymmdd” or “mmddyy.” In a

Figure 1-5-1. An entry for DC element *date*.

Term Name: date	
URI:	http://purl.org/dc/elements/1.1/date
Label:	Date
Definition:	A point or period of time associated with an event in the lifecycle of the resource.
Comment:	Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
References:	[W3CDTF] http://www.w3.org/TR/NOTE-datetime

Source: DC 1.1: http://dublincore.org/documents/dces/, accessed 2007.

metadata instance embedded in a Web page, the <meta name="DC.date" scheme="iso8601" content="2006-05-03" /> tag indicates that the value of *date* element follows ISO 8601 (Data elements and interchange formats—Information interchange—Representation of dates and times) and should be read as 2006 May 3 rather than 2006 March 5. Another example is the *subject* element, which in many standards requires a controlled vocabulary or classification scheme to constrain the number and source of values. If the record has assigned terms from the LCSH, the subject element may use the scheme attribute to tell users that the subject headings come from LCSH. These rules and vocabularies are designated *value encoding schemes*.

The guidance provided in metadata standards assists with the correct implementation of principles, and offers best practices guidelines for localizing the usage in both the implementation and creation of metadata records. This assists metadata creators in making the right choices about which data values as well as what order, syntax, and form that the data values should be entered into any record. It also reflects the best practices that have been culled from particular user communities. Best practices may vary greatly in scope and level of detail. Some standards provide guidance in a user manual or recommendations, while others offer well-established subject vocabularies, name authorities, or term lists developed by an authoritative source. A higher level of guidance would be to provide rules, recommendations, and examples for selecting, ordering, and formatting data used to key selected elements. Data contents guidelines may be found in the documentation of element sets in most metadata specifications, and sometimes monographic publications, e.g., *Cataloging Cultural Objects* (CCO).

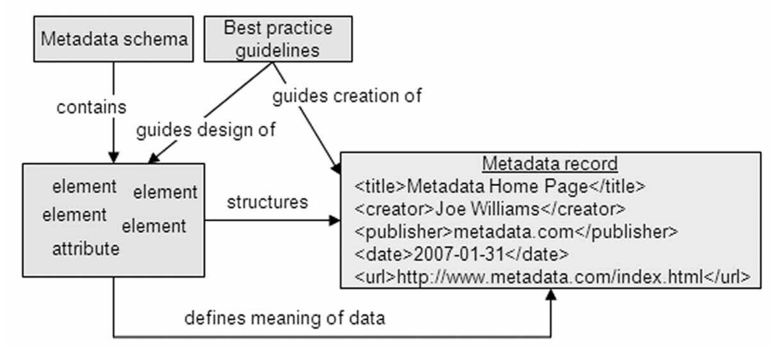
Defined element sets are often represented in XML and the resultant structures are known as *metadata schemas*. These representations define what an element means and how it will be encoded in a machine-processable format.

Figure 1-5-2 provides an illustration of metadata components. A metadata schema provides a framework for structuring data in records, while the labels of elements indicate function or define the data. User manuals or best practices guidelines ensure a sound technical (and consistent) metadata element set, as well as the correct application of a set when creating records. The metadata record example in Figure 1-5-2 presents four of the DC metadata elements (*title*, *creator*, *publisher*, and *date*). Chapters 3 and 4 provide further explanation.

Certain standards may maintain both a simple version that contains only basics and a full version with additional element *refinements* (see DC examples, Chapter 2). Standards that employ attributes to modify and refine elements also need to maintain these sets of attributes and their allowable values (see MODS examples, Chapter 2). With an increase in XML applications, metadata standards developers now often provide a version of the element sets expressed in an XML schema, e.g., CDWA's offspring: CDWA Lite. Many developers have also created crosswalks to map their own elements with the elements of other related standards.

In the next chapter selected metadata standards are introduced, with the focus on their *semantics* rather than syntax or how metadata records should be

Figure 1-5-2. Relationships between a metadata schema and a record.



created using these standards. Those topics are detailed in Chapters 3, 4, and 5, which cover the structure and semantics of metadata schemas, the syntax of schema encoding, and the issues and practices in generating metadata records, respectively. Chapters 2 through 5 serve as building blocks to design metadata schemas and create metadata records. Chapters 6 through 8 cover more advanced topics concerning metadata services, interoperability, and evaluation. Chapter 9 discusses research developments of the past thirteen years, and touches on trends of the near future.

Suggested Readings

Caplan, Priscilla. 2003. "Metadata Basics." In *Metadata Fundamentals for All Librarians*. Chicago: American Library Association: 1–11.

Gilliland, Anne J. 2000. "Setting the Stage." In *Introduction to Metadata: Pathways to Digital Information*, edited by Murtha Baca. Online Edition (Version 2.1). Los Angeles, CA: Getty Information Institute. Online Edition Version 2.1. Available: www.getty.edu/research/conducting_research/standards/intrometadata/setting.html.

Jul, Erik. 1995. OCLC Internet Cataloging Project. *D-Lib Magazine* [Online], December 1995. Available: www.dlib.org/dlib/december95/briefings/12oclc.html.

NISO. 2004. *Understanding Metadata*. Bethesda, MD: NISO Press. Available: www.niso.org/publications/press/UnderstandingMetadata.pdf.

Weibel, Stuart, Jean Godby, Eric Miller, and Ron Daniel. 1995. *OCLC/NCSA Metadata Workshop Report*. Available: http://dublincore.org/workshops/dc1/report.shtml.