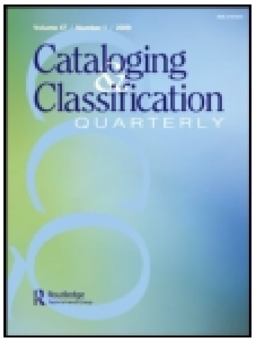


Cataloging & Classification Quarterly

ISSN: 0163-9374 (Print) 1544-4554 (Online) Journal homepage: https://www.tandfonline.com/loi/wccq20

Understanding Metadata and Metadata Schemes

Jane Greenberg

To cite this article: Jane Greenberg (2005) Understanding Metadata and Metadata Schemes, Cataloging & Classification Quarterly, 40:3-4, 17-36, DOI: 10.1300/J104v40n03_02

To link to this article: https://doi.org/10.1300/J104v40n03_02

Published online: 23 Oct 2009.

PART I: INTELLECTUAL FOUNDATIONS

Understanding Metadata and Metadata Schemes

Jane Greenberg

SUMMARY. Although the development and implementation of metadata schemes over the last decade has been extensive, research examining the sum of these activities is limited. This limitation is likely due to the massive scope of the topic. A framework is needed to study the full extent of, and functionalities supported by, metadata schemes. Metadata schemes developed for information resources are analyzed. To begin, the author presents a review of the definition of metadata, metadata functions, and several metadata typologies. Next, a conceptualization for metadata schemes is presented. The emphasis is on semantic container-like metadata schemes (data structures). The last part of this paper introduces the MODAL (Metadata Objectives and principles, Domains, and Architectural

Jane Greenberg is Associate Professor, School of Information and Library Science, University of North Carolina at Chapel Hill, 205 Manning Hall, CB #3360, Chapel Hill, NC (E-mail: janeg@ils.unc.edu).

[Haworth co-indexing entry note]: "Understanding Metadata and Metadata Schemes." Greenberg, Jane. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 40, No. 3/4, 2005, pp. 17-36; and: *Metadata: A Cataloger's Primer* (ed: Richard P. Smiraglia) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2005, pp. 17-36. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

Available online at http://www.haworthpress.com/web/CCQ

doi:10.1300/J104v40n03_02

Layout) framework as an approach for studying metadata schemes. The paper concludes with a brief discussion on the value of frameworks for examining metadata schemes, including different types of metadata schemes.

KEYWORDS. Information resources, metadata, metadata schemes, data structures, MODAL (Metadata Objectives and principles, Domains, and Architectural Layout) framework

INTRODUCTION

Digital repositories have grown at an explosive rate over the last decade due to new information technologies, particularly those supporting World Wide Web (Web) applications. This growth has led to a tremendous increase in the need for data management, an intense interest in metadata in a wide range of communities (e.g., education, government, scientific, business, etc.), and extensive development of metadata schemes. There are hundreds of metadata schemes being used, many of which are in their second, third, or *n*th iteration. Many specifications developed for information resources have been endorsed by standards bodies. For example, the Dublin Core Metadata Element Set, Version 1.1 (2003) (hereafter referred to as the Dublin Core) has been formally endorsed as a standard by Committeé Européen De Normalisation (CEN) as CEN Workshop Agreement (CWA) 13874 (http://www.cenorm.be/cenorm/businessdomains/businessdomains/informationsocietystandardizationsystem/published+cwas/13874.pdf), the National Information Standards Organization (NISO) as NISO Z39.85-2001 (http://www.niso.org/standards/resources/Z39-85.pdf), and most recently the International Standards Organization (ISO) as ISO 15836-2003 (http://www.niso.org/international/SC4/n515.pdf).

An official list of all available metadata schemes does not exist, not even one specific to information resources, although a number of metadata registries are becoming fairly extensive. The CORES registry (http://www.cores-eu.net/registry/) is a good example, currently listing 40 metadata schemes, and supporting searching and browsing by metadata scheme developer, maintenance agency, element sets, elements, encoding schemes, application profiles, and element usages. Additionally, Web resources, such as the IFLA DIGITAL LIBRARIES: Metadata Resources Web page (http://www.ifla.org/II/metadata.htm) and the UKOLN Metadata Web site (http://www.ukoln.ac.uk/metadata/), provide ample information about metadata schemes by referencing and linking to papers, presentations, electronic listservs and newsgroups, and metadata specifications. These resources are useful for studying the population of metadata schemes, although it is difficult to study this topic in its entirety, given the multiplicity of schemes, their evolutionary nature (different versions or releases produced over time), their different constituencies, and their varied functional emphases.

Despite these challenges, it is important to conceptualize the nature of metadata schemes. We need to study schemes in order to understand their place in the larger context of information organization, management, and access. The intense interest in, and extensive development of, metadata schemes over the last decade brings this need to the forefront and calls for an examination of the population of these semantic container-like systems. A framework is needed to study the full extent of and functionalities supported by metadata schemes. In this paper I consider this need by examining schemes developed for information resources. The first part is a review of the definition of metadata, metadata functions, and several metadata typologies. Next, I present conceptualization for metadata schemes. The emphasis is on semantic container-like metadata schemes (data structures). The last part of this paper introduces the MODAL (**M**etadata **O**bjectives and principles, **D**omain, and **A**rchitectural **L**ayout) framework as an approach for studying metadata schemes. The paper concludes with a brief discussion on value of frameworks for examining metadata schemes, including different types of metadata schemes.

METADATA AND METADATA SUPPORTED FUNCTIONS

Defining Metadata

It is well documented that Jack E. Myers coined the term "metadata" in 1969, and it was first printed in a product brochure in 1973. Meyers used the term to represent current and future products associated with his MetaModel and to designate a company that would develop and market those products. METADATA® was registered in 1986 as a U.S.

Trademark for The Metadata Company, where Meyers is a principle (The Metadata Company: http://www.metadata.com/).

Building upon Meyer's usage, the terms *metadata*, *meta data* and *meta-data* (all in lower case) have been adopted by the computer science, statistical, database, and library and information science communities to mean "data about data." The term *metadata*, in these realms, addresses data attributes that describe, provide context, indicate the quality, or document other object (or data) characteristics.

Information and library scientists both equate (e.g., Milstead & Feldman, 1999; Caplan, 1995) and distinguish (e.g., Heery, 1996) creating metadata from cataloging. The main distinction is that metadata is exclusive to electronic information. This interpretation is, however, not hard-and-fast, as librarians have been cataloging electronic resources for decades prior to the Web, and many metadata schemes are applicable to physical as well as digital resources. Definitions of metadata specific to information resources (the types of materials found in both physical and digital libraries, archives, museums, and other information agencies) are consistent in that they emphasize the functional aspect of metadata, with the common definition of "structured data about data" (e.g., Duval et al. 2002; Woodley et al. 2003). Metadata can be viewed as "structured data about an object that supports functions associated with the designated object"–with an object being "any entity, form, or mode for which contextual data can be recorded" (Greenberg, 2003).

Metadata Functions

Many discussions, particularly those exploring metadata in the information resource community (libraries, archives, museums, and other information centers), tend to group metadata elements by the various functions they support. The result is the identification of different types of metadata (or metadata classes), each of which comprises multiple metadata elements. Table 1 provides typologies of different types of metadata identified by Lagoze et al. (1996), Gilliland-Swetland (2000), Greenberg (2001), and Caplan (2003).

Lagoze et al. have developed one of the most extensive typologies, presented in Table 1, columns one through three. Column one summarizes the Lagoze et al. typology, column 2 describes the metadata functions corresponding to the typology, and column three lists examples of metadata elements that facilitate the functions in column 2.

Gilliland-Swetland's (2000), Greenberg's (2001), and Caplan's (2003) typologies are presented in Table 1, columns four, five, and six respec-

TABLE 1. Metadata Typologies and Functionalities

Lagoze et al. (1996):			**Gilliland-Swetland (2000): Typology of 5 types of metadata**	**Greenberg (2001): Typology of 4 types of metadata (2 sub-types of Use metadata)**	**Caplan (2001): Typology of 4 types of metadata**
Typology of 7 types of metadata	**Metadata Functions "This type of metadata *facilitates*":**	**Element examples***			
Identification/ description metadata	***RESOURCE DISCOVERY/ INFORMATION RETRIEVAL***	Creator (Author), Title, Subject	Descriptive metadata	Discovery metadata	Descriptive metadata
Administrative metadata	***RESOURCE MANAGEMENT***	Price, Condition	Administrative and Preservation metadata	Administrative metadata	Administrative metadata
Terms and conditions metadata	***RESOURCE USAGE***	Rights, Reproduction restrictions	Administrative, Preservation, and Use metadata	Technical Use, Intellectual Use, and Administrative metadata	Administrative and linking metadata
Content ratings metadata	***RESOURCE USE BY APPROPRIATE AUDIENCES***	Audience	Use metadata	Technical Use and Intellectual Use metadata	Administrative and linking metadata
Provenance metadata	***RESOURCE AUTHENTICATION AND OTHER PROVENANCE-RELATED ACTIVITIES***	Creator, Source	Administrative and Use metadata	Authenticity and Administrative metadata	Administrative metadata
Linkage/ relationship metadata	***RESOURCE LINKING WITH RELATED RESOURCES***	Relation, Source	Administrative metadata	Authenticity and Administrative metadata	Linking metadata
Structural metadata	***RESOURCE HARDWARE AND SOFTWARE NEEDS***	Compression ratio	Technical and Use metadata	Technical Use metadata	Structural metadata

*Individual metadata elements can be multi-functional. For example, "source" metadata facilitates resource authentication and resource linking, and can be classed as both "Provenance metadata" and "Linkage/relationship metadata" following Lagoze et al.'s typology (for a discussion on metadata element multifunctionality see Greenberg, J. (2001)).

tively; and these typologies are mapped to the functions identified by Lagoze et al. (column 2). The typologies developed by these three authors (Gilliland-Swetland, Greenberg, and Caplan) are not as extensive as the Lagoze et al. typology and definitions vary among authors. As a result, metadata types are repeated in the mapping, and often more than one type is listed to match the Lagoze et al. metadata functions.

Table 1 illustrates similarities among metadata typologies, yet it also makes evident the challenge in establishing one universal metadata classification. Most important is that the naming of different types of metadata, with labels such as "resource discovery" and "use," demonstrates that *functionality* is the principal reason for metadata.

A CONCEPTUALIZATION FOR METADATA SCHEMES

A Context for Metadata Schemes

While literature includes analyses on the different types of metadata, discussions on the meaning of *scheme* (or *schema*), in the metadata context, are scarce. It may simply be that "scheme" is generally understood to be "a structured framework or plan" (*Miriam-Webster Online*: http://www.m-w.com/cgi-bin/dictionary?va=schema), and adopting this concept for representation systems used in information databases seems very reasonable. The term *scheme* has historically been applied to classificatory and terminological systems used in library catalogs and other information databases, such as the *Dewey Decimal Classification* (*DDC*) system and *Library of Congress Subject Headings* (*LCSH*); this practice continues with little debate–if any. Here I focus on the semantic container-like schemes, also referred to as data structures. These are the higher-level structured schemes that may require or recommend the use of schemes containing acceptable data values (e.g., *DDC* or *LCSH*).

The philosopher Kant provides insight into the meaning of *scheme* in his 1785 treatise *Critique of Pure Reason* (1998). Kant reasons that a *schema* is a system based on experience and the gathering of empirical data. Kant's model, emphasizing experience and empirical analysis, is applicable to developments underlying metadata schemes today. The experience aspect is evident via expert collaboration during scheme design activities. Subject discipline experts and technical experts, in areas such as electronic markup, thesaurus development, representation, and data processing, frequently join forces through committees or designated initiatives to design a metadata scheme. These committees and initiatives can be locally, nationally, or internationally positioned. For example, the Dublin Core Metadata Initiative is an international and interdisciplinary group that has developed and maintains the Dublin Core through committee consensus. Metadata scheme designers, working together, draw upon experience gained in their areas of expertise.

Referring back to Kant's model, the empirical aspect of scheme design is evident mainly through the counting of metadata elements, sub-elements, qualifiers, and other components. Nascent metadata scheme design can be guided by a pre-determined number of elements, dictating the scheme's extent and often granularity (refinement). This restriction requires element counting during the design process. The empirical aspect of metadata scheme design is also evident in analyses of pre-existing metadata schemes–even those developed during the pre-digital era. The crosswalk analysis, a common analytical method, often employed as a first step to scheme design, is heavily empirical (e.g., Woodley, 2000; Zeng, 1997). Equivalent or similarly functioning metadata elements from two or more metadata schemes are mapped to one another via a table that allows for an element-level detailed comparison among schemes. Other empirical methods include counting how many institutions or initiatives have adopted a scheme; counting the number of metadata elements being used (or not used) in a particular scheme; and analyzing the strength of the relationship among metadata elements (e.g., Moen and Benardino, 2003).

The "conceptualization" of a metadata scheme is generally "formalized in a specification" (Greenberg, 2003), and there are standards to guide this process. Arguably, the most important standard guiding this process is ISO/IEC (International Standards Organization/International Electrotechnical Commission) 11179, Metadata Registries Standard, developed by ISO/IEC JTC1 SC32 WG2 Development/Maintenance (http://metadata-stds.org/11179/). The standard has six parts: (1) Framework, (2) Classification, (3) Registry Metamodel and Basic Attributes, (4) Formulation of Data Definitions, (5) Naming and Identification Principles, and (6) Registration. The standard includes extensive instructions on how to identify data elements and register a scheme with a registration authority. ISO/IEC 11179 is essential for the database community, and has been a vital resource for the development of metadata schemes for digital resources. Part 1 and Part 4 of ISO/IEC 1179 appear to be the most important sections for metadata scheme development. Part 1 "introduces and discusses fundamental ideas of data elements, value domains, data element concepts, conceptual domains, and classification schemes," and Part 4 "provides guidance on how to develop unambiguous data definitions." Additional standards guiding metadata scheme development include ISO/IEC 20943, Procedures for Achieving Metadata Registry Content Consistency; ISO/IEC 20944, Metadata Registry Interoperability and Bindings; and ISO/IEC 18038, Identification and Mapping of Various Categories of Jurisdictional Domains (see http://metadata-stds.org/).

Metadata Scheme Conceptualized

Exploring the meaning of "scheme," including referencing ISO/IEC 11179, and reviewing metadata functionalities aid in conceptualizing what a metadata scheme is. Described as "a systematic, orderly combination of elements or terms" (Woodley et al., 2003) and "a set of metadata elements and rules for their uses that has been defined for a particular purpose" (Caplan, 2003), literature does not reveal a universally accepted definition for metadata scheme–unlike the standard definition of "data about data" for metadata. A metadata scheme can, however, be identified by three main features. A metadata scheme is:

1. A collection of metadata elements gathered to support a function, or a series of functions (e.g., resource discovery, administration, use, etc.), for an information object.
2. A collection of metadata elements, forming a structured container, to which data values are added. Data values may be uncontrolled or controlled (e.g., taken from a source such as *LCSH* or a standardized list of values).
3. A collection of data elements, with their attributes formalized in a specification (or a data dictionary). Examples of element attributes include the metadata element's "name," "identifier," "label," "definition," and the "date the element was declared."

Specifications vary tremendously from fairly flexible guidelines, such as the Dublin Core, to detailed and complex rules, such as Federal Geographic Documentation Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) (1998), or the CSDGM's extension for biological data (Content Standard for Digital Geospatial Metadata Part 1: Biological Data Profile, 1999). Specifications almost always provide element definitions–first and foremost. In reference to other element attributes, specifications might or might not:

- Define relationships among metadata elements.
- Identify specific acceptable content values or content value systems (e.g., *LCSH*) to be used with the scheme.
- Provide syntactical guidelines. Content syntax guidelines, such as "surname" before "forename" or "year" "month" "day" [YYYY-MM-DD], or encoding and markup syntax guidelines for XML or HTML languages.
- Declare metadata element *cardinality*. That is, how many times a metadata element can or must appear in a metadata description.
- Further refine metadata elements definitions through qualification. For example, the metadata element "creator" can be further refined by "person" or "corporate body."

ANALYZING METADATA SCHEMES

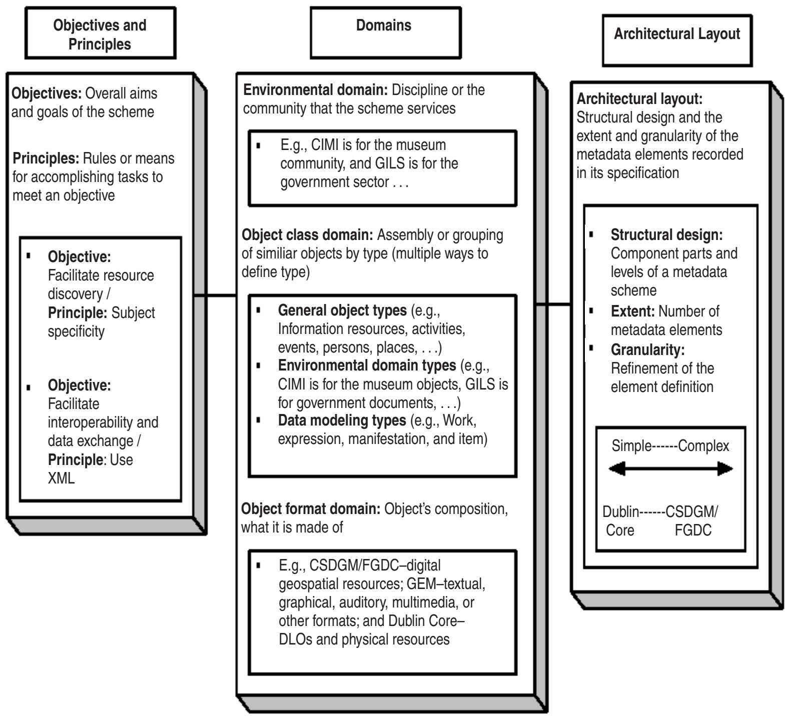
A comprehensive study of metadata schemes seems daunting when considering the wide range of metadata schemes, their evolutionary nature (different versions or releases produced over time), their different constituencies, and their varied functional emphases. Despite this challenge, research is needed to better understand the range of metadata schemes and how they fit into the information context. The MODAL framework, defined by *objectives and principles*, *domains*, and *architectural layout*, provides a way to examine the population of metadata schemes. Figure 1 illustrates the MODAL framework, which is discussed in more detail in this section.

Objectives and Principles

One way to further understand a metadata scheme is to study its underlying objectives and principles. Objectives identify the overall aims and goals of the scheme, while principles are rules or means for accomplishing tasks to meet an objective. Objectives and principles are often published as a series of statements, and frequently appear as metadata specification introductions. These underlying features of metadata schemes are often inter-woven, making it difficult to distinguish the objectives from the principles.

Among the most influential historical statements of bibliographic control objectives, are Charles A. Cutter's (1904) objectives for a library catalog, printed in the 4th edition of his *Rules for a Dictionary Catalog*. Emphasizing resource discovery, almost one hundred years

FIGURE 1. The MODAL Framework for Metadata Objectives and Principles, Domains, and Architectural Layout



prior to the development of the World Wide Web and digital technologies, Cutter's objectives state that a library catalog is to:

1. Enable a person to find a book when the author, title, or subject is known;
2. Show what the library has by author, subject, and literature genre; and
3. Assist in the selection of a book by its edition and literary or topical composition. These objectives have a functional end: they exemplify resource discovery, and with a little tweaking of terminology (e.g., changing book to digital resource), they represent the objectives of many current metadata schemes developed for digital resources.

Cutter's *Rules for a Dictionary Catalog* (1904) also includes principles to meet the stated objectives. An example of the subject specificity principle, which aids Cutter's first two objectives, is: "Enter a work under its subject-heading, not under the heading of a class which includes that subject. Example: Put Lady Cust's book on *The cat* under **Cat**, not under **Zoology** or **Mammals**, or **Domestic animals**; and put Garnier's *Le fer* under **Iron**, not under **Metals** or **Metallurgy**" (1904, p. 66).

Among one of the best known historical examples of principles of bibliographic control is the "Statement of Principles," commonly known as the "Paris Principles," resulting from the International Conference on Cataloguing that took place in Paris in 1961 (International Federation . . . , 1963). Adopted by delegates from 53 different countries, the Paris Principles were established to guide the selection and form of access points for library materials and to harmonize and direct future cataloging on an international level (Carpenter, 1985). It is interesting to note that the second statement in the Paris Principles is a rendition of Cutter's objectives.

The synthesis of objectives, principles, and functional requirements continues to be a fundamental component of metadata schemes that guide online cataloging. Exploring these underlying components was a major theme in the 1996 International Conference on the Principles and Future Development of *AACR* (Weihs, 1998), which explored the adaptation of *AACR2* for cataloging electronic resources. The *Functional Requirements for Bibliographic Records* (*FRBR*) (1997), an International Federation of Library Associations and Institutions' (IFLA) model for bibliographic control, provides another recent example of a synthesis of these fundamental components.

Historical and contemporary cataloging developments are indicative of the objectives and principles stated in metadata schemes developed for digital resources. A key objective of many metadata schemes is to facilitate resource discovery, and, thus many specifications indicate this. The first item listed in the Dublin Core Metadata Initiative's mission statement Web page (http://www.dublincore.org/about/) is to improve means for finding resources on the Internet by "developing metadata standards for discovery across domains." Another objective, found in a number of fairly recent metadata schemes, is to document preservation and reformatting. The final report of RLG's Working Group on Preservation Issues of Metadata (1998) states that the scheme's 16 elements are deemed "crucial to the continued viability of a digital master file." This objective, or variations of it, is found in the introductions of many other preservation metadata schemes (for links to preservation metadata see: http://www.dublincore.org/groups/preservation/).

A number of metadata schemes have formal statements of principles to help accomplish objectives and to guide scheme design. A case in point is the Poughkeepsie Principles, an outcome of the Vassar Planning Conference that led to the Text Encoding Initiative (TEI) guidelines (Design Principles . . . , 1988). These Principles consist of nine statements that highlight the need to establish a set of guidelines for data interchange, define a meta-language, recommend encoding principles, support compatibility with existing standards, and permit the conversion of resources to newer communication formats, among other goals–all for electronic text in the humanities. Many of these statements appear more like objectives and weave in the principles. The TEI standard includes a series of metadata schemes for different types of text in the humanities, such as novels, plays, lexicons, and bibliographic information–to name a few components. The most current version is TEI P4 (Sperberg-McQueen and Burnard, 2002).

Another set of metadata scheme principles is the Ann Arbor Accords (1996) (hereafter referred to as the Accords), developed for metadata about electronic archival finding aids. The Accords were influenced by the Poughkeepsie Principles, and appear to also interweave objectives and principles. The Accords "defines principles and criteria for designing, developing, and maintaining SGML-based encoding schemes for archive and library finding aids," and have guided the development of the SGML (now XML) document type definition (DTD) for the Encoded Archival Description (EAD) (2002). The "General Principles" section of the Accords (principles 4 through 8) defines the chief aim of archival finding aids, and emphasizes goals of the EAD, such as facilitating interchange and portability, increasing the intelligibility of finding aids within and across institutions, and fostering data sharing. The "Structural Features" section of the Accords (principles 9 through 12) includes EAD design principles. This section explains that the EAD encoding scheme is based on SGML with a formal DTD, that the scheme consists of two parts (an SGML-compliant DTD and detailed application guidelines), and that more generic terms, such as "unit" and "component," were used to ensure broad application of the scheme.

A final example of principles with a slightly different tenor than the Poughkeepsie Principles and the Ann Arbor Accords are the Dublin Core's principles of simplicity, semantic interoperability, international consensus, and flexibility for Web resource description (e.g., Dublin

Core Metadata, 1997). These principles define general modes of operation and are the tenets upon which the Dublin Core has developed.

Objectives and principles, as reviewed here, highlight one way in which the population of metadata schemes may be examined and further understood. In the next section, I will focus on metadata scheme domains.

Domains

Additional insight into the population of metadata schemes can be gained by examining a scheme's application domains–that is, the realms in which the scheme operates. This discussion presents three domains for studying metadata schemes. They are the environmental domain, object class domain, and object format domain. The distinctions between these domains are not absolute, but they are discussed separately for the sake of clarity.

Environmental Domain

The environmental domain is the discipline or the community that the scheme serves. The majority of schemes in the Web environment have been developed to meet the needs of a specified community. Examples include Consortium for the Computer Interchange of Museum Information (CIMI) (2000), developed primarily for the museum community and art world; Government Information Locator Service (GILS) (1997), developed for the government sector; Rich Site Summary (RSS) (2000), developed for the news community; and Data Documentation Initiative (DDI) (2000), developed for the social science research community. The Dublin Core is a multidisciplinary schema developed to facilitate resource discovery and support interoperability and data exchange among communities. As a result, the Dublin Core's environmental domain includes the larger information resource description community (Weibel, 1995), which may be thought of as a community uniting smaller communities or disciplines.

Object Class Domain

Object class domain is the assembly or grouping of similar objects by "type." There are a variety of ways that type can be defined for the object class domain. A majority of the schemes identified in this article (e.g., Dublin Core, TEI, and EAD) have been designed for information resources–the class of objects commonly found in libraries, museums, archives, or other information agencies. Other types of objects include activities, events, persons, places, structures, relationships, transactions, execution directions, and programmatic applications. Each of these object types can have sub-domains, which can be further divided by genre. For example, information resources can be further refined by format (e.g., textual, graphical, auditory, and multi-media, among other formats), a topic that is covered below under *object format domain*.

The environmental domains identified above each cater to a specific type of resource, and illustrate another way to look at the object class. For example, CIMI is for art objects, GILS is for government documents, RSS is for channel information, and DDI is for social science statistical data.

Data modeling activities lend to another way of typing objects. For example, the *FRBR* model supports representation of works, expressions, manifestations, and items, which may be viewed as different classes or objects. Traditionally, metadata schemes have combined these representations. The *Machine Readable Cataloging* (*MARC*) bibliographic format illustrates this point and is considered a flat data structure. Research on the nature of a work and multiple derivations (e.g., Smiraglia, 2001) is important for forwarding information object representation, particularly in the digital environment where derivations are likely to be even more significant compared to the high numbers already found in catalogs representing more traditional bibliographic resources (e.g., monographs, videos, musical works, and so forth) (Smiraglia and Leazer, 1999).

Object Format Domain

Object format domain is the object's composition–what it is made of. FGDC/CSDGM was designed for digital geospatial materials and is an example of a format-specific scheme. The majority of information resource-oriented metadata schemas are applicable to multiple formats. For example, the Gateway to Educational Material (GEM) Element Set (1999) is a metadata schema for educational resources, available in textual, graphical, auditory, multimedia, or other formats. The Dublin Core, designed initially for digital like objects (DLOs), which are defined as textual objects, is also applicable to physical objects in many different formats (Weibel and Hakala, 1998). Extending beyond information resources, object format is a likely factor in other schemas as well. Consider the ontology developments in the field of genomics, where metadata schemes designed for human beings or animals document different blood types, e.g., the DNA (deoxyribonucleic acid) gene sequence. This example is beyond the scope of this article, but it demonstrates another view of domain format.

Architectural Layout

Architectural layout involves the structural design of a scheme, and the extent and granularity of the metadata elements. (Scheme architecture is not to be confused with information architecture, which applies to design framework for entire information systems or environments.) Scheme structures range from fairly simple flat designs–that generally have a limited number of general metadata elements–to complex modular structures with a higher number and usually more granular metadata elements.

At the simple end of the spectrum is the Dublin Core, which comprises 15 basic metadata elements that are deemed essential for resource discovery. The simplicity of the Dublin Core is predicated on its existence as a low-level common denominator that aims to support interoperability among resource description environments using more complex schemas. This activity is facilitated by crosswalk analyses that map metadata elements to achieve semantic interoperability. Another factor underlying the Dublin Core's simplicity is connected to its goal to serve as the core or base for the development of more elaborate or domain specific schemas. An example is found with GEM, which incorporates all 15 Dublin Core metadata elements and their exact definitions, but includes a set of specialized metadata elements (e.g., "pedagogy," "grade level," "audience," to name a few) that are important for organizing and accessing Web-based educational resources. Resources containing GEM metadata can easily be transferred to any database that uses equivalent Dublin Core elements.

Slightly more complex than the Dublin Core's architecture, but still relatively simple, is Metadata Object Description Schema (MODS) (2003). This scheme has 19 top level elements and two root elements, each with sub-elements and most with attributes to describe the element. Another slightly more complex metadata scheme is the Visual Resource Association's Core Categories (VRA Core, 3.0) (2000) for visual objects (e.g., paintings, antiques, other cultural objects) and images that document them (e.g., slides, photographs, digital images used for studying these objects). Initially modular, with one component for the work (the visual object) and one component for the image of the work (see Core Categories for Visual Resources, 1997), this schema now conflates these two parts into a single arrangement. VRA, version 3.0, has 17 metadata elements and recommends one metadata record per object, whether the work itself or an image of the work is being represented. The VRA is more granular than the Dublin Core in that the elements are specifically defined for the visual resource domain.

Compared to the VRA Core, the DIG35 Specification: Metadata for Digital Images (2000) has a fairly complex architectural layout. This scheme includes five top-level structural components: basic image, image creation, content description, history, and intellectual property rights, and over 150 metadata elements for documenting images. The "image creation" section includes sub-sections labeled general information, camera capture metadata, device characterization, camera settings, scanner capture metadata, and captured item metadata, each of which are further defined by a collection of metadata and meta-metadata elements, the majority of which are not prescribed in the VRA Core.

A final example of a scheme with a complex architectural layout is the FGDC/CSDGM developed for digital geospatial resources. The structural design consists of ten top-level sections; this schema has over 320 compound metadata elements. FGDC/CSDGM, Section 4, Spatial Reference Information, provides a good example of this scheme's granularity with over 40 metadata elements–all of which map to the Dublin Core "coverage" metadata element to achieve semantic interoperability. For example, a data exchange sequence between the two schemes would require that FGDC/CSDGM "Map_Projection_Parameters," "Grid_Coordinate_System," "Universal_Transverse_Mercator," "State_Plane_Coordinate_System," "ARC_Coordinate_System," "Planar_Coordinate_Information," "Distance_and_Bearing_Representation," and "Geodetic_Model" metadata elements (a few element examples) be mapped to the Dublin Core's "coverage" metadata element.

In observing schema architectures, particularly the extent of and granularity of metadata elements, factors such as metadata element cardinality and qualification must also be taken into account. This is because element repeatability in one specification might be able to achieve the extensity and granularity documented in another specification. This point is illustrated with the example above where FGDC/CSDGM, Section 4, metadata elements, which are quite as granular and have obligation and occurrence restrictions, were mapped to the Dublin Core "coverage" metadata element, which is optional and repeatable.

Although the architectural layout examples provided here are fairly representative of scheme architectural compositions, it is important to point out that greater complexity in structural design is not always correlated with a higher number of and more granular metadata elements. For example, RSS has a slightly complex structural design with four main components (channel, item, text input, and image), but only 16 metadata elements that, with the exception of the granular descriptions given for the "channel," are applicable to a wide range of information resources. The EAD scheme for archival finding aids deviates from the norm as well, albeit in the other direction. EAD has a simple structure comprising two main components (header and archival description), but is extensive in that it contains over 143 metadata elements with descriptions that range from general to quite granular.

CONCLUSIONS: FRAMEWORKS AS AN APPROACH OF STUDY

Metadata is generally defined as structured data about data, and information and library scientists emphasize the functional goals in their discussions. The functional aspect is evident when reviewing metadata typologies. Defining what specifically a metadata scheme is presents a greater challenge, given the multiplicity of schemes, their evolutionary nature, their different constituencies, and their varied functional emphases. The MODAL framework presents a way to study this complex topic.

Frameworks are useful for understanding complex topics: they help divide, categorize, and analyze concepts. Although metadata schemes vary tremendously, they are shown to be similar when examining their objectives and principles, domain foci, and architectural layout. The MODAL framework focuses on these features and provides a structure for examining and interpreting metadata schemes. The MODAL framework can also aid with metadata scheme design.

Although frameworks facilitate study, it is important to note that they are artificial creations, with accompanying shortcomings. The MODAL framework will likely not support every analysis of metadata schemes, and it might require enhancement and modification over time. It is also likely that researchers will develop new frameworks supporting analyses of different types of metadata schemes. This point becomes evident when considering the different uses of the term metadata scheme: "metadata scheme" connotes not only the semantic container-like structures focused on in this article, but also topics we commonly refer to as content standards (e.g., *AACR*), controlled vocabularies (e.g., *LCSH*), taxonomies and ontologies (e.g., Yahoos! main classification), and full markup language, such as Mathematical Markup Language (MathML™) (1999). These different types of metadata schemes require further investigation, and point to an exciting research agenda.

Studying more semantic container-like metadata schemes, as emphasized in this paper, and the different types of metadata schemes noted above will help improve the MODAL framework approach. The MODAL framework can be used to study different types of metadata schemes, and testing of its applicability is warranted. Together these approaches will contribute to a body of knowledge about the population of metadata schemes and help us understand the role of metadata schemes in the larger context of information organization, management, and access.

WORKS CITED

Ann Arbor Accords: Principles and Criteria for an SGML Document Type Definition (DTD) for Finding Aids. (1996). http://sunsite.berkeley.edu/FindingAids/EAD/accords.html.

Caplan, P. 1995. You call it corn, we call it syntax–Independent metadata for document-like-objects. *The Public-Access Computer Systems Review*, 6 (94): http://info.lib.uh.edu/pr/v6/n4/capl6n4.html/.

Caplan, P. 2003. *Metadata Fundamentals for All Libraries*. Chicago: American Library Association.

Carpenter, M. 1985. Statement of the Principles: International Conference on Cataloguing Principles, Paris, October, 1961, Editor's Introduction. In Svenonius, E. and Carpenter, M. (Eds), *Foundations of Cataloging: A Source Book*. Littleton, CO: Libraries Unlimited, p. 176-178.

CIMI Consortium. 2001. CIMI Consortium for the Computer Interchange of Museum Information. Guide to Best Practice: Dublin Core, Version 1.1. http://www.cimi.org/old_site/documents/meta_bestprac_v1_1_210400.pdf.

Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile. 1999. Biological Data Working Group Federal Geographic Data Committee and USGS Biological Resources Division. http://www.fgdc.gov/standards/documents/standards/biodata/biodatap.html.

Cutter, C. A. 1904. *Rules for a Dictionary Catalog*, 4th ed., (rewritten). Washington, DC: Government Printing Office.

Data Documentation Initiative (DDI) Codebook DTD. 2000. http://www.icpsr.umich.edu/DDI/CODEBOOK.TXT.

Design Principles for Text Encoding Guidelines: TEI ED P1. 1988 [rev. January 1990]. (The Poughkeepsie Principles were original published November, 13, 1987. The original publication is not accessible on the Web.)

DIG35 Specification: Metadata for Digital Images, Version 1.0. 2000. Digital Imaging Group, Inc. http://www.bgbm.fu-berlin.de/TDWG/acc/Documents/DIG35-v1.0-Sept00.pdf.

Dublin Core Metadata Element Set, Version 1.1: Reference Description. 2003. http://www.dublincore.org/documents/2003/06/02/dces/.

Duval, E., Hodgins, W., Sutton, S., and Weibel, S. 2002. Metadata Principles and Practicalities. *D-Lib Magazine* http://www.dlib.org/dlib/april02/weibel/04weibel.html.

Encoded Archival Description Tag Library, *Version 2002*. 2002. Prepared and Maintained by the Encoded Archival Description Working Group of the Society of American Archivists and the Network Development and MARC Standards Office of the Library of Congress.

Federal Geographic Metadata Committee. 1998. *Content Standard for Digital Geospatial Metadata* (*CSDGM*). http://fgdc.er.usgs.gov/metadata/csdgm/.

Functional Requirements for Bibliographic Records [final report]. 1997. Recommended by the IFLA Study Group on the Functional Requirements for Bibliographic Records; International Federation of Library Associations and Institutions, IFLA Universal Bibliographic Control and International MARC Programme. Frankfurt Am Main: IFLA UBCIM. http://www.ifla.org/VII/s13/frbr/frbr.htm.

Gateway to Educational Materials (GEM) Element Set. 1999. http://www.geminfo.org/Workbench/Metadata/GEM_Element_List.html [last modified 2001].

Gilliland-Swetland, A. 2000. Defining metadata, in M. Baca (ed.), *Introduction to Metadata: Pathways to Digital Information*. Los Angles, CA: Getty Information Institute. http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/index.html.

Government Information Locator Service (*GILS*). 1997. http://www.gils.net/prof_v2.html.

Greenberg, J. 2001. A quantitative categorical analysis of metadata elements in image applicable metadata schemas. *Journal of the American Society for Information Science and Technology* 52: 917-914.

Greenberg, J. 2003. Metadata and the World Wide Web. *Encyclopedia of Library and Information Science*, pp. 1876-1888. New York: Marcel Dekker, Inc.

Heery, R. 1996. Review of metadata formats. *Program*, 30: 345-373.

International Federation of Library Associations and Institutions. 1963. *Report: International Conference on Cataloguing Principles, Paris, 9th-18th October, 1961*. London: Organizing Committee of the International Conference on Cataloguing Principles, p. 91-96.

Kant, Immanuel. 1998. *Critique of Pure Reason*. Paul Guyer and Allen W. Wood, eds., and translators. Cambridge: Cambridge University Press.

Lagoze, C., Lynch, C. A., and Daniel, R. 1996. The Warwick Framework: A container architecture for aggregating sets of metadata. Available at: http://cs-tr.cs.cornell.edu:80/Dients/Repository/2.0/Body/ncstrl.cornell%2fTR96-1593/html.

Mathematical Markup Language (MathML™) 1.01 Specification W3C Recommendation. 1999. Available at: http://www.w3.org/TR/REC-MathML/.

Milstead, J. and Feldman, S. 1999. Metadata: Cataloging by any other name. *Online* 25-31. Also available at: http://www.onlineinc.com/onlinemag/OL1999/milstead1.html.

MODS Version 3.0. 2003. http://www.loc.gov/standards/mods/v3/mods-3-0.xsd.
Moen, W. E., and Benardino, P. 2003. Assessing metadata utilization: An analysis of *MARC* content designation use. In Sutton, S., Greenberg, J., and Tennis, J. eds. *2003 Dublin Core Conference: Supporting Communities of Discourse and Practice–Metadata Research and Applications. DC-2003: Proceedings of the International DCMI Conference and Workshop. September 28-October 2, 2003, Seattle, Washington*. Syracuse, NY: Information Institute of Syracuse.
RDF Site Summary 1.0 Specification (*RSS*). 2000. http://www.egroups.com/files/rss-dev/specification.html#.
RLG and Preservation. 1998. Working Group on Preservation Issues of Metadata Final Report. http://www.rlg.org./preserv/presmeta.html.
Smiraglia, R. P. 2001. *The Nature of 'A Work': Implications for the Organization of Knowledge*. Lanham, MD: Scarecrow.
Smiraglia, R. P. and Leazer, G. H. 1999. Derivative bibliographic relationships: the work relationship in a global bibliographic database. *Journal of the American Society for Information Science* 50: 493-504.
Sperberg-McQueen, C. M. and Burnard, L. eds. 2002. *TEI P4: Guidelines for Electronic Text Encoding and Interchange*. Text Encoding Initiative Consortium. XML Version: Oxford, Providence, Charlottesville, Bergen.
Visual Resource Association Core Categories for Visual Resources (VRA CORE), version 2.0 1997. http://www.oberlin.edu/~art/vra/wc1.html.
VRA Core Categories, Version 3.0. 2000. Visual Resources Association Data Standard Committee. http://www.vraweb.org/vracore3.htm [last modified 2002].
Weibel, S. and Hakala, J. 1998. A report on the workshop and subsequent developments. *D-Lib Magazine*: http://www.dlib.org/dlib/february98/02weibel.html.
Weibel, Stuart. 1995. Metadata: The foundations of resource description. *D-Lib Magazine*. http://www.dlib.org/dlib/July95/07weibel.html.
Weihs, J. ed. 1998. *The Principles and Future of* AACR2*: Proceedings of the International Conference on the Principles and Future Development of* AACR2. Chicago: American Library Association.
Woodley, M. 2000. Metadata standards crosswalks. In Baca, M. ed., *Introduction to Metadata: Pathways to Digital Information*. Los Angeles, CA: Getty Information Institute. http://www.getty.edu/research/conducting_research/standards/intrometadata/3_crosswalks/index.html.
Woodley, M. S., Clement, G. and Winn, P. 2003. *DCMI Glossary*. http://dublincore.org/documents/2003/08/26/usageguide/glossary.shtml [last update, April 2004].
Zeng, M. L. 1999. Metadata elements for object description and representation: A case report from a digitized historical fashion collection project. *Journal of the American Society for Information Science* 50: 1193-1208.