Metadata and Authority Control

Sherry L. Vellucci

A variety of information communities have developed metadata schemes to meet the needs of their own users. The ability of libraries to incorporate and use multiple metadata schemes in current library systems will depend on the compatibility of imported data with existing catalog data. Authority control will play an important role in metadata interoperability. In this article, I discuss factors for successful authority control in current library catalogs, which include operation in a well-defined and bounded universe, application of principles and standard practices to access point creation, reference to authoritative lists, and bibliographic record creation by highly trained individuals. Metadata characteristics and environmental models are examined and the likelihood of successful authority control is explored for a variety of metadata environments.

As the Internet becomes an accepted source of electronic information, librarians and information specialists around the world strive to improve methods for description, organization, and retrieval of remotely accessed documents and other electronic objects. They are not alone in this endeavor, for within many other sectors of our society the creators and providers of electronic resources are also engaged in attempts to manage this vast body of information. This has resulted in the simultaneous and parallel development of a wide variety of metadata schemes. Each group has approached the problem of organization and access from its own frame of reference, and included in its metadata model the type of information that best serves the needs of its own user community. It remains to be seen whether this proliferation of metadata schemes will improve description and access for library users, or only compound current problems. Much will depend on the ease with which metadata can be incorporated into current library systems and the compatibility of imported metadata with existing online catalog data. In other words, the successful use of multiple metadata schemes in the library environment will depend on authority control.

In this article, I offer a starting point for future discussions and research on metadata and authority control within the library profession. I provide an overview of metadata and discuss a few of the more popular metadata schemes associated currently with libraries, museums, archives, and information centers. These are examined within the context of authority control and its role in the future metadata environment.

Metadata: What Are They?

The most common definition of metadata is "data about data." While accurate, this definition is so simplistic that it adds very little to our understanding of the concept. Further elaboration of the origins and use of the word will provide a framework for this discussion. Early use of the term "metadata" goes back to the

Sherry L. Vellucci (velluccs@stjohns.edu) is Associate Professor, St. John's University Division of Library and Information Science, Jamaica, New York.

1960s, but it began to appear more frequently in the literature on database management systems (DBMS) in the 1980s. The term "metadata" was used to describe the information that documented the characteristics of information contained within databases (Phillips 1995). In this DBMS domain, the computer was the setting for both the information being described and the descriptive data, and, therefore, metadata operated within a totally electronic environment.

The parallel world of library cataloging traditionally used the terms "bibliographic data" or "cataloging data" for this type of information. Catalogers used these terms when both the objects they cataloged and the bibliographic records were in a nonelectronic form. Catalogers continued to call this type of information cataloging data when the bibliographic record migrated to the MAchine Readable Cataloging (MARC) formats. They even called it bibliographic or cataloging data when they began to organize and describe local computer files. But when catalogers began to describe *networked* electronic resources using the same type of bibliographic data, the terminology changed. Suddenly the MARC record became metadata, and the cataloger's familiar world—circumscribed by the Anglo-American Cataloguing Rules (AACR) and the MARC formats—changed forever. This transformation was caused by a convergence with the broader world of information organization. The methods of organizing resources from the rather separate domains of library science, computer science, and information science all converged in this networked environment, and the term "metadata" became the commonly accepted term in all disciplines.

Some people in the library cataloging community balk at the term, noting that metadata are merely cataloging or bibliographic data by another name (Larsgaard 1996). This is true, of course, but there were some very good reasons for adopting this new terminology. The term metadata already had the connotation with an electronic environment that was lacking with the term cataloging; but more importantly, as catalogers expanded their own world view of information organization, metadata provided a common term to use when communicating with other information organizers in the computer and information science world. The term allowed everyone to interact equally in the continuum of information organization. This equality is important for the development of bibliographic systems that access, transfer, and manipulate metadata information across multiple databases in a way that provides meaningful information to the user. If, however, metadata are focused solely on electronic resources, there is danger of segmenting access to the wide spectrum of resources at the very time everyone is striving to make the continuum of information organization as seamless as possible. A broader perspective of both resources and information organization must be adopted by all constituencies in order to avoid perpetuating this type of segmentation, and the term metadata should be applied to data for all types of resources.

In order to place metadata within its proper context, I use the following expanded definition, which incorporates statements of functionality and environment: Metadata are data that describe the attributes of a resource; characterize its relationships; support its discovery, management, and effective use; and exist in an electronic environment. While this definition reflects the metadata ideal, in reality, most metadata schemes do not accomplish all of these functions equally well. Few metadata sets express relationships adequately. Many schemes concentrate more on resource discovery and retrieval than on detailed descriptions that act as surrogates for the object. This quickly becomes evident when comparing some metadata element sets to the data provided in a typical catalog record.

Before examining specific metadata schemes in more detail, it is important to understand the fundamental principles of authority control and why it is generally successful in our current catalog environment. Once the authority control success factors are understood, each metadata scheme can be analyzed to predict whether authority control will operate successfully when applied to that environment.

Authority Control in the Library Environment

Volumes have been written about the role of authority control in the online environment, but it is only recently that the concept of *online* was extended beyond the local online catalog. Taylor (1994) was among the first to discuss the issue of authority control in the broader context of the Internet. To date there has been little research in this area, and the question remains: Will the basic principles and processes of authority control work with metadata in the broader networked environment? Before we can answer this question, we must understand why authority control has worked so successfully in library catalogs.

Traditionally, the purpose of authority control has been to bring consistency to library catalogs. The authority control process is directed at the access points contained in catalog records, i.e., names, titles, and subjects. It ensures that these access points are unique and consistent in content and form, and provides a network of linkages for variant and related headings in the catalog. Reference to an authority file maintains consistency in controlled access points, while cross-references and the adjacent display of identical access points have been the primary methods used to accomplish the necessary linkages in the catalog. In Web-based catalogs, however, hypertext links are beginning to provide more direct catalog linkages.

Authority control is closely allied with the functions of the catalog. These functions, which were first defined by Cutter more than a century ago, continue to be the starting point for most discussions of authority control (Younger 1995). They are commonly described as finding, collocating, evaluating, and locating. Recently, the International Federation of Library Associations and Institutions (IFLA) Study Group on the Functional Requirements for Bibliographic Records (FRBR) translated these functions into "user tasks," as part of an entity-relationship modeling exercise. The FRBR model suggests that catalogs should support the following four basic user tasks (IFLA 1998, 90):

1. To find entities that correspond to the user's stated search criteria (i.e., to locate either a single entity or a set of entities in a file or database as the result of a search using an attribute or relationship of the entity)
2. To identify an entity (i.e., to confirm that the entity described corresponds to the entity sought, or to distinguish between two or more entities with similar characteristics)
3. To select an entity that is appropriate to the user's needs (i.e., to choose an entity that meets the user's requirements with respect to content, physical format, etc., or to reject an entity as being inappropriate to the user's needs)
4. To acquire or obtain access to the entity described (i.e., to acquire an entity through purchase, loan, etc., or to access an entity electronically through an online connection to a remote computer)

Authority control plays a major role in helping the user successfully complete most of these tasks. It supports the finding task by ensuring that each entity has a unique name; that only one name is used for each entity; that variant name forms are represented and linked in some way; and that related names, titles, and subjects are collocated in the catalog. Authority control supports the identifying function by enabling users to distinguish among entities with similar names, titles, or subjects through the use of unique and consistent headings. While it can't be said that authority control directly supports the selecting function, which depends on a wide variety of individual user needs, the access points provided by authority control could be influential in the user's selection of a particular entity.

The final user task—obtaining access to the entity— is achieved by using the call number as a location device or shelf address. Classification as a location device works well for physical items found in local collections; however, with remotely accessed electronic resources, the means of "physically" accessing the object must be an accurate identifier of some type. At present, this identifier is most often the Uniform Resource Locator (URL)—an address notori-

ous for its instability—that is now embedded in the bibliographic record. The success of the URL as a means of accessing the entity depends on accuracy in this location information. As a working solution, OCLC developed an authority control system called the Persistent Uniform Resource Locator (PURL) for handling these changing addresses (Jul 1997). In addition, work progresses on other naming devices such as the Digital Object Identifier (DOI) and the Uniform Resource Name (URN), and these may become the primary means of location identification. Interestingly, inclusion of this type of identifier in the bibliographic record does not currently fall within the rubric of the cataloging rules. While the USMARC format was revised to accommodate these locators, similar revision has not yet been made to AACR. The closest rule that exists in the latest version of AACR is for the standard number (i.e., ISBN or ISSN), which, although used as an identifier, is not a location address. It is likely, however, that when stable identifiers are developed for networked electronic resources and a standard is adopted by the networking community, these will become a part of the authority control process in some way.

Authority Control Success Factors

There are several reasons for the success of authority control, beginning with the fact that it operates within a well-defined and bounded universe—the library catalog. These catalogs should be viewed as value-added systems that describe and provide access to a highly selective, quality-controlled collection. A second reason for the success of authority control is that the creation of access points is based on principles and standardized practices that guide the process, including AACR and *LC Rule Interpretations*. In addition, authority work is aided by reference to authoritative lists such as the *Library of Congress Name Authority File*, the *Library of Congress Subject Headings*, *Sear's Subject Headings*, or the *Art & Architecture Thesaurus*, to name a few. And finally, the cataloging process, and therefore authority work, is performed by highly trained individuals who are part of a library culture and understand cause and effect in the information retrieval process.

All of these factors contribute not only to the integrity of local library catalogs, but also to the success of shared cataloging on a national level. Cataloging within an international arena, however, presents new authority control challenges. Several authors and projects have identified difficulties that arise because of geographic, language, and cultural bias when attempting to implement international authority control and international bibliographic data transfer (Borgman 1997; Bourdon 1994; Delsey 1989; Murtomaa and Grieg 1994; *REUSE* 1997; Tillett 1996; Vellucci 1998). Borgman (1997) suggests that to operate

effectively in the international arena, each library must think globally while acting locally. This advice should also apply to the creation and use of metadata, but the metadata environment is even more complex because of the multiple schemes and structures involved. So, what are the possibilities for authority control within the global context of the metadata environment?

Authority Control in the Metadata Environment

I should begin by saying that there is no single metadata environment—which, of course, is the problem. Anything discussed here might apply to some metadata schemes, but not all. For example, the MARC bibliographic record is a metadata scheme to which authority control is applied regularly in terms of its data content and use. But this is the scheme employed most often in the controlled library environment; what happens when we go beyond MARC? What other types of schemes exist? What is the culture of these metadata communities? Is authority control feasible or likely in metadata schemes beyond MARC?

Metadata Schemes and Architectures

Many metadata schemes are currently in use or proposed for use within an information setting. The following list provides a sampling of the more well-known metadata sets:

- Computer Interchange of Museum Information (CIMI)
- Federal Geographic Data Committee (FGDC) Content Standards for Digital Geospatial Metadata
- Dublin Core (DC) Metadata Element Set
- EDUCOM Instructional Management Systems
- Encoded Archival Description (EAD)
- Government Information Locator Service (GILS)
- IAFA Templates (IAFA/WHOIS++)
- USMARC Formats
- Resource Description Framework (RDF)
- Text Encoding Initiative (TEI) Header
- Visual Resources Association (VRA) Core Data

Some of these metadata schemes are general in nature—such as the MARC format or the DC—and are designed to accommodate information about electronic resources in a wide variety of disciplines. Other metadata schemes are more specialized—such as CIMI and the FGDC Content Standards for Digital Geospatial Metadata—and apply to digital information in a specific format or within a specific discipline or domain. What they all have in common is that they contain a set of defined data elements that describe the

entity and help provide access to it in the electronic environment. Beyond that, they all vary as to the number of data elements, the content of the data elements, and the standards used, if any, for that content. In addition, some of these metadata schemes extend beyond the traditional functions of description and access to include resource management and use information such as terms and conditions data, administrative data, content ratings data, provenance data, linkage or relationship data, and structural data.

Metadata Characteristics

There are three basic characteristics common to all metadata schemes: (1) syntax, (2) semantics (i.e., content), and (3) structure. A scheme's syntax can range from a highly complex format, such as the MARC record or the SGML-encoded TEI header, to a basically unstructured scheme, such as the original DC. The semantics can include scores of complex data elements, whose content is prescribed by standards and rules, or it can have as little as two or three elements with no control at all over that content. Metadata can be contained in a variety of database structures (or architectures), including library catalogs, commercial database packages, or the recently formalized RDF standard. With increased importance placed on global access to information and system interoperability, many of the new metadata schemes avoid the more complex syntax and rigid semantic content prescribed by library cataloging rules, MARC formats, controlled vocabularies, and traditional classification schemes. In some cases, however, metadata users and creators want more from their metadata than a simple structure can offer. As a result, some schemes are being modified to include more elements, more data qualifiers, and a more complex structure. In addition, a few metadata sets (e.g., the DC and EAD) have added tagging conventions to support the use of authoritative data for names and subjects, and to indicate the authoritative sources for controlled headings.

Metadata Buzzwords

Flexibility, interoperability, and extensibility are the new buzzwords in digital information organization, and each of these concepts will influence, or be influenced by, authority control of metadata. Flexibility allows the metadata creator to include as much or as little detail as desired in the metadata record, with or without adherence to any specific cataloging rules or authoritative lists. Flexibility in the amount of metadata detail does not present problems for authority control. The various levels of description prescribed by AACR and the Program for Cooperative Cataloging core records already demonstrate this. Flexibility with regard to rules and authoritative lists for data content

and form, however, would make authority control very difficult indeed.

Interoperability, as defined by Borgman, refers to "the ability of information systems to interact in a useful way on a real-time basis over communications networks" (Borgman 1997, 231). For the sake of interoperability, many believe that the metadata record should contain a basic core set of data elements that would be common to all schemes. This would facilitate the exchange and use of metadata in a variety of systems and is the basic premise of the DC. Crosswalks offer a way to map the data elements from one metadata scheme to another and are a major part of the interoperability factor. Several crosswalks currently exist to map the MARC format to and from other popular metadata schemes such as the DC and GILS (Day 1997). These crosswalks focus on transfer architecture and syntax, however, and do not address interoperability in terms of the standards used for data content and form. Thus, while it would appear that authority controlled data would improve interoperability, it is not a requirement for the physical transfer of the data. Authority control becomes an interoperability issue only when metadata from various sources and in various formats need to be integrated and manipulated for efficient use.

The final buzzword—extensibility—means that the scheme should allow the addition (extension) of data elements and data qualifiers to accommodate specific user needs (i.e., be domain specific). The danger with extensible data is a lessening of interoperability, because the more the general metadata elements are qualified or extended on an individual or domain specific basis, the less compatible or interoperable the metadata become. Extensibility, however, is not necessarily a problem for authority control. In fact, many of the data qualifiers being added to metadata schemes are designed to identify the rules or authoritative lists used to control the content of a given field, and, as such, they enhance metadata authority control.

Standard Generalized Markup Language-Based Metadata

Just as the MARC format provides the encoded record syntax for most online catalogs today, markup languages form the encoded syntax for many documents found on the Internet (Gaynor 1996). Unlike MARC, however, where the bibliographic record is separate from the document that it describes, markup languages can be used to encode both the descriptive data and the object. This allows metadata to be embedded within the document itself. Standard Generalized Markup Language (SGML) is one of the most common markup languages used for metadata, along with its related codes HyperText Markup Language (HTML), and the

newer Extensible Markup Language (XML). The following sections examine two of the more popular metadata schemes that use SGML: TEI headers and EAD.

TEI Headers

One of the earliest metadata schemes to use SGML was developed within the scholarly communities of literature and linguistics. A common encoding scheme was needed for complex textual documents in order to facilitate the exchange and use of the growing corpus of electronic texts in the humanities. The TEI developed guidelines for the preparation and interchange of electronic texts in the broadest sense of the term, including both textual and nontextual resources in electronic form (Sperberg-McQueen and Burnard 1994). The TEI also developed guidelines for descriptive documentation for each TEI encoded document. This required documentation is called the TEI header, and much of its data content is similar to title page information.

TEI headers consist of four components that provide details about the text, its source, its encoding practices, and its revision history: the File Description (demarked by the tags <fileDesc> and </fileDesc>); the Encoding Description (with <encodingDesc> and </encodingDesc>); the Profile Description (<profileDesc> and </profileDesc>); and the Revision Description (<revisionDesc> and </revisionDesc>). This TEI header information can be embedded within the TEI-conformant text that it describes (such as header information in HTML encoded Web documents), or it can be maintained in a separate file and point to the electronic text (such as the bibliographic records in a library catalog).

The File Description is the only required component of the four TEI Header segments. AACR cataloging standards were used as a model for this portion of the header. The elements that make up the File Description (and which roughly correspond to the areas of International Standard Bibliographic Description [ISBD]), consist of the following components:

- Title Statement:
  <titleStmt> . . . </titleStmt>
- Edition Statement:
  <editionStmt> . . . </editionStmt>
- Extent:
  <extent> . . . </extent>
- Publication Statement:
  <publicationStmt> . . . </publicationStmt>
- Series Statement:
  <seriesStmt> . . . </seriesStmt>
- Notes Statement:
  <notesStmt> . . . </notesStmt>
- Source Description:
  <sourceDesc> . . . </sourceDesc>

Three data elements are mandatory in the File Description segment: the Title Statement, which includes both title and author data; the Publication Statement; and the Source Description. The Profile Description segment of the header, which describes the content of the document rather than the document itself, contains the data elements for controlled vocabularies.

Just as there are various levels for description in the cataloging world, the amount of data in a TEI header can range from very simple to extremely complex, depending on the nature of the text, its intended use, and the creator of the metadata. In general, TEI headers are created by the person who created the full electronic text; in other words, people who might have little knowledge of authority control or the concept of uniform name headings. Some institutions, such as the University of Virginia, are applying content standards to this metadata—that is, ISBD, AACR, and authority control for names and subjects (Shieh 1998). This is an institutional policy decision, however, and requires cataloger intervention to review and revise the header data after creation; these content standards are recommended but not mandated by TEI guidelines (Sperberg-McQueen and Burnard 1994). Currently, there is no official TEI-MARC crosswalk to map information from the TEI Header to the MARC record, although the University of Virginia developed and uses its own model.

The Encoded Archival Description

Another SGML-based metadata scheme was developed more recently within the archival community. Motivated by the increased role of networks in accessing information and a desire to expand on the data provided in the MARC record, a cooperative project based at the University of California, Berkeley developed a nonproprietary SGML encoding standard for machine-readable finding aids—the EAD (Thibodeau, Pitti, and Barry 1996). EAD was designed as a data structure and communication standard for the inventories and registers created by archives, libraries, museums, and manuscript repositories to support the use of their holdings. Such finding aids typically describe a unique collection, its intellectual organization, and its components; therefore, finding aids are metadata in their own right. The EAD finding aid document consists of three segments: the Header (<eadheader> . . . </eadheader>); Front Matter (<frontmatter> . . . </frontmatter>) ; and the Finding Aid (<findaid> . . . </findaid>).

The developers of EAD borrowed several concepts from the TEI, including the use of a header to contain information about the rest of the encoded document—in this case, the finding aid. Thus, the EAD Header is meta-metadata, that is, it is information that describes descriptive information. The header is required by the Document Type Definition (Library of Congress 1997), and provides information that describes the finding aid itself, such as might be found on an "electronic title page." The EAD header consists of the following five data segments:

- EAD Identifier:
  <eadid> . . . </eadid>
- File Description:
  <filedesc> . . . </filedesc>
- Profile Description:
  <profiledesc> . . . </profiledesc>
- Revision Description:
  <revisiondesc> . . . </revisiondesc>
- Footer:
  <footer> . . . </footer>

Like the TEI header, the File Description segment of the EAD header provides the basic bibliographic data, but these data are about the finding aid, not the collection being described. Data describing the collection, which catalogers normally think of as bibliographic data, are contained in the Archival Description element (<archdesc . . . </archdesc>) of the Finding Aid portion of the document.

As with many metadata schemes, EAD defines structure and content designation. Like the TEI guidelines, it recommends the use of content standards for collection description. EAD differs from other metadata types in that it accommodates multi-level, hierarchical descriptions that are necessary in the archival environment. Typically, MARC records for archival materials provide collection level descriptions, while finding aids might describe at the collection, series, file, or item levels. The information at the collection level, therefore, is most often mapped to the MARC record, and it is at this content level that the archival community frequently applies the recommended cataloging standards of AACR and the related cataloging guidelines for *Archives, Personal Papers, and Manuscripts* (Hensen 1989).

Encoding tags for controlled access points are an important authority control aspect of EAD. These tags allow the authoritative heading and its source to be identified. From its inception, the developers of the EAD displayed a high level of awareness of basic cataloging standards, and while adherence to these standards is not mandated, it is strongly recommended. Thus, the EAD metadata structure was designed to accommodate standards and, therefore, authority control, if desired.

The Dublin Core Element Set

Fast becoming the metadata scheme of choice for general library use, the DC was named for the site of the first Metadata Workshop held in Dublin, Ohio. It is maintained by OCLC and was developed within an international and

interdisciplinary environment through the collaboration of various Internet communities. The DC defines a set of 15 basic data elements for resource description and retrieval (*Dublin Core* 1997). The 15 elements, many of which correspond to data in traditional catalog records, are divided among three information categories: Content (Title, Subject, Description, Source, Language, Relation, Coverage), Intellectual Property (Creator, Publisher, Contributor, Rights), and Instantiation (Date, Type, Format, Identifier) (Hillman, 1998).

Designed for simplicity and flexibility, the DC is not governed by mandatory rules. All elements are optional, all elements are repeatable, the order of elements is optional, and all elements can be qualified. Included among the optional qualifiers is the qualifier "scheme," which identifies the authority or standard that is used for that particular metadata element's content. Initially, the DC was designed to be syntax independent; that is, no rules of grammar were mandated for the data so that the data elements could be used in many different environments and with any type of database architecture. But implementation of DC projects required a syntax that could be used consistently on the Web without changes to existing Web standards and software. HTML <META> tags were adopted by many early projects as the preferred syntax, and are currently in the process of becoming a formalized standard; however, since the Resource Description Framework's (RDF) approval by the World-Wide-Web Consortium (W3C) as a structural standard, XML has become the recommended syntax for use with the more sophisticated RDF architecture (Miller 1998).

The DC was originally conceived for author-generated description of general Web resources, and several metadata authoring tools based on the DC have been designed for a nonlibrary market. Tools such as Meta Tag Builder use forms to prompt Web page developers to include DC metadata for their Web documents at the time of document creation. In this context, DC metadata is intended to provide useful information to enhance indexing by Internet search engines. Other authoring tools such as DC Dot will automatically seek out an existing Web page and generate the DC HTML <META> tags. There is little likelihood, however, that Web page creators who generate metadata outside the library and information environment will have any understanding of authority control; and while many of the DC data elements such as Title, Creator, and Subject correspond to areas presently under authority control in library catalog records, the DC does not mandate that data form or content conform to any particular rules or guidelines.

Figures 1 and 2 demonstrate the flexibility of the DC in terms of content. Figure 1 shows an example of some basic DC data elements with HTML <META> tags. A Web-site developer who wanted to provide data for Internet search engines to index created these metadata, which contain free text data and are technically correct by DC standards.

Figure 2 shows DC metadata for the same Web site, created by a cataloger according to cataloging standards and conventions, for eventual conversion to a MARC record. It is also technically correct by DC standards.

These DC examples show that a given metadata's *accommodation* of standards for data elements is not enough to ensure successful authority control. It is a necessary first step, without which the remaining discussion is pointless, but it is only one part of a complex system. Ultimately, effective authority control also depends on the environment in which the metadata are created and used.

The Metadata Environment and Culture

Weibel, the driving force behind the DC, offers three different models for the metadata environment (Weibel 1997). The first is the "embedded model," where metadata is formulated by the creator of the electronic document and resides within the resource being described using the HTML <META> tag. In this first model, the metadata are integral to the resource, require no additional organizational system to be used, and can be harvested by Web indexing agencies. This is an increasingly popular Internet metadata model, as the general Internet user population becomes aware of metadata and search engines are reconfigured to exploit the HTML <META> tags.

The second model is the "third party metadata model," where an agency creates, collects, and manages separate metadata records that refer to resources but are not embedded in the resource. This model requires an organizational system to use the metadata and is exemplified by bibliographic records in a library catalog. This second model might link the separate metadata records to their

<META NAME = "DC.title" Content = "Welcome to My
   Metadata Home Page">
<META NAME = "DC.creator" Content = "Ima D. Expert">
<META NAME = "DC.subject" Content = "Hot Metadata
   Sites">

Figure 1. Dublin Core Uncontrolled Content

<META NAME = "DC.title" Content = "Metadata Home
   Page">
<META NAME = "DC.creator" Content = "Expert, Ima D.,
   1975-">
<META NAME = "DC.subject" Content = "Metadata">

Figure 2. Dublin Core Controlled Content

electronic resource using the HTML <HREF> tag and the URL address.

Weibel's third model is the "view filter model," where a distinct agency manages many metadata records from a variety of sources that use different metadata structures and maps these variant metadata sets into a common set such as the DC. This third model provides users with a single query construct and facilitates a cross-disciplinary network infrastructure. It is dependent on a metadata architecture such as the Warwick Framework (Lagoze1996) or the RDF (World Wide Web Consortium 1997) to facilitate the data transfer and mapping. SOLINET's Community Planet project is one example of this model.

The four authority control success factors discussed earlier (controlled operating environment, trained provider, application of standards, and reference to authoritative lists) correlate to several factors in Weibel's models of the metadata environment. All three models address the issues of where the metadata reside and how they are used. These factors map to the first authority control success factor: controlled operating environment. At issue for authority control, however, is not so much whether the metadata are embedded within the document or maintained as separate records, but rather what type of "organizational system," to use Weibel's phrase, is used to exploit the metadata. Indeed, this is one of the most important factors in successful authority control. If the organizational system is designed to implement the controlled vocabulary, uniform access points, and syndetic structure created by the authority control process, then authority control can flourish in the metadata environment. The organizational systems of Weibel's second and third models can be so designed. If, however, there is no organizational system, as in the first model, or if the organizational system is not designed to take full advantage of standardized metadata content, then authority control can be only partially successful at best. Thus, the operating environment is critical to the success of authority control.

A second factor in Weibel's metadata models addresses the issue of who provides the metadata, which will determine whether the metadata content is standardized. These issues map to two success factors: the trained provider and the application of standards. In some cases, the author of the document will provide the metadata, especially when the metadata are embedded or contained within the document. Although there has been some attempt to ensure that a few select data elements are included in metadata created by the use of authoring tools such as Meta Tag Builder (which uses DC.Lite—a pared down version of the DC element set), there is little likelihood that Web-page creators in general will apply content standards to metadata. This lack of standardized metadata content in the majority of Web sites, while creating difficulties for the average Web surfer, is beyond the control of the library and information community (just as much information is not collected and cataloged

by libraries today). Only when these metadata are brought into an organizational system within its purview can the library community successfully apply standards to the data.

Templates also were developed to encode documents and create headers for several metadata schemes used within the information community, including the TEI, EAD, FGDC, and DC. Often developed for a particular project, these templates prompt the creator to include a specific data element, but, again, this will not guarantee that content standards for authority control of names, titles, or subjects will be applied. There is, however, greater awareness of the existence of cataloging standards among those who use these metadata schemes, and, therefore, the likelihood of standardized metadata content may be greater within these metadata communities than it would be among Web site creators in the general population.

Information specialists and catalogers will create metadata for only a small percentage of electronic resources, concentrating primarily on high quality and long-lasting documents. In these cases, authority controlled data content should be the norm. It is likely, however, that a larger body of bibliographic records will be derived from a variety of existing metadata; in fact, I predict that this will become a major part of the future copy cataloging process. Currently, the existing metadata must be reviewed by catalogers to ensure adherence to authority control standards before integrating these metadata records into authority controlled catalogs. Authority control productivity tools, such as Northwestern's *Toolkit* (Strawn, 1995), could further automate this process, and the potential exists in future environments, such as Weibel's third model, for the organizational system architecture to process existing authority data automatically.

This prospect leads to the last authority control success factor—reference to authoritative lists. While this factor is not overtly discussed in any of Weibel's models, it is an implicit component of both the second and third models. Certainly, reference to authoritative lists is part of the authority control process for library cataloging exemplified in the second model; but Weibel's third model also will depend to a great extent on this type of referencing, although these authoritative lists most probably will be called data registries. In the RDF metadata model these authoritative lists or data registries provide meta-metadata that can be used: (1) to identify the metadata type from among various schemes; (2) to verify and validate a particular metadata element's semantics and syntax; (3) to verify the agency that created the metadata; (4) to verify specific standards used for data content and form; (5) to identify specific resources; and (6) to exchange data elements among metadata packages. These authority links are necessary not only because of the variety of metadata schemes, but also because of the many different content standards and subject lists used in various disciplines. Weibel's third model, there-

fore, will require extensive building of authoritative data registries and extensive relationship modeling.

Finally, the various metadata environments can be viewed as cultures within which the metadata creators operate. The value system imposed on information organization within each culture will determine the extent to which authority control succeeds in terms of data integration for library systems. These cultures can be grouped into three broad sections: the library cataloging culture, the scholarly information culture, and the popular information culture.

In the library cataloging culture, authority control flourishes. Metadata providers are highly trained professionals whose primary job is the creation of metadata. They are not only aware of the need for authority control, but practice it on a daily basis, especially when functioning in a networked environment that includes shared bibliographic data. Until recently, authority controlled metadata in this culture was almost exclusively in the form of MARC records; but as awareness and understanding of other metadata schemes increases, catalogers are eager to extend the benefits of their authority control knowledge to other systems.

The scholarly information culture consists of many different groups, defined primarily by discipline, each with its own metadata scheme (e.g., TEI, EAD, FGDC, CIMI, VRA, etc.). Often, those who developed the metadata schemes were aware of the concept of authority control and accommodated it to some extent in the metadata's syntax. In this culture, metadata creators might be concerned with the creation of information as well as its organization, and while they frequently operate in close proximity to the library cataloging culture, usually they are not part of it. Authority control may exist, but often uses standards that differ from those in the library cataloging culture. Increased communication between the library cataloging culture and the scholarly information culture resulted in the development of crosswalks and data conversion processes for the exchange of metadata between these two cultures.

The popular information culture is primarily concerned with creating and sharing information, rather than organizing it. In this culture, metadata creators have almost no contact with the other two cultures and are not aware of authority control concepts. HTML <META> tags, when used, generally contain natural language and uncontrolled content. It is highly unlikely that content standards will be applied in this culture and a great deal of work is required for the library cataloging culture to make use of any metadata created in the popular information culture.

Beyond MARC: Future Possibilities

The potential for authority controlled metadata certainly exists; however, it will not be universally applied in all sec-

tors of the metadata environment. Multiple metadata schemes can be used in the library environment in a variety of ways, but the transfer and conversion of these metadata records into bibliographic records for use in current library catalogs requires both accuracy in structural mapping from one system to another and data compatibility in terms of content and form. On the surface, crosswalks appear to provide a simple means of exchanging data. For example, the DC Title element maps to the 245 field in USMARC; few things in the world of metadata, however, are as simple as they seem. No metadata crosswalks work equally well in both directions. Mapping 15 DC metadata elements into MARC may not seem difficult, but as additional element qualifiers are implemented to improve resource description, the mapping difficulty increases. In addition, mapping a standard MARC record into 15 simple Dublin Core elements presents a real challenge. Much of the data-rich MARC record is lost in the process. To date, the effective use of crosswalks for data exchange has been only partially successful at best.

The metadata community is looking in another direction for solutions to the problems of structural mapping and data compatibility. Recent activities of the DC Metadata Initiative indicate that several areas currently targeted for improvement may eventually solve some of these problems (Weibel 1999). These activities include: (1) formalization of the DC maintenance process; (2) formal standardization of the element set; (3) mechanisms for addition to data qualifiers, and (4) implementation of the RDF.

The DC developers recently completed an organizational restructuring that provides an explicit process for decision making. In addition to policy and technical advisory committees, the new structure provides for working groups to address particular problems with the metadata scheme. Similar to other agencies such as the W3C or the Internet Engineering Task Force, the new directorate created a formal process for ratifying changes to the DC with input from various communities. Prior to the next stage of formalization, which is submission to both NISO (National Information Standards Organization) and CEN (Center for European Normalization) for status as an official standard, the data element definitions will be reviewed to improve clarity and promote more consistent application. Additionally, the data elements themselves will be formatted according to ISO 11179, a standard for formally expressing the semantics of data elements. Formalization of the standard and standardization of the element set is a critical step that will provide stability for structural mapping.

Data element qualification is another area targeted for improvement. While the basic 15 elements of the DC have remained stable for three years now, virtually all DC applications have adopted element qualifiers to accommodate better precision in resource description than the basic

model supports. This highlights the paradox that exists in the DC community: on one hand, the more elements and element qualifiers that are added, the less interoperable the data become; on the other hand, no one is really satisfied with the quality and level of description afforded by the basic element set. The ability to add standard qualifiers to the DC elements would improve the scheme, especially in the area of authority control. Standard qualifiers would: (1) increase semantic specificity by allowing controlled vocabularies or classification schemes to be identified and used for either description or as a formal browsing structure; (2) allow authority controlled data to identify uniquely a name, title, work, object, etc.; and (3) allow definition of a formal substructure so that compound values can be assigned to one element. Provision of formal qualifiers that identify standards used to control data content and form should allow greater specificity without reducing interoperability. The RDF will play a critical role in this solution, for although the qualifiers will provide improved precision, it is the data structure that allows specific data values to be clearly identified and associated with specific metadata schemes, functions, roles, or actions.

The RDF

The RDF is a set of conventions for expressing metadata that uses XML as an encoding standard and provides a framework for exchanging metadata. A Universal Resource Identifier (URI) is used for naming purposes, and each resource has properties that describe it. By using the XML-namespace facility, RDF associates each property with its associated metadata schema that defines the property, thus allowing data elements from different metadata schemes to reside together without confusion or conflict. The ability to specify metadata schemas in RDF makes it possible for applications to access a particular scheme from a publicly accessible registry on the Web and retrieve the parsing structure and semantics of the element set. In addition, this data model allows elements to reference data registries (i.e., authority files), thus providing a sophisticated version of authority control. Figure 3 shows an example of one possible RDF record construct.

If the DC is moving in the direction of formalization, standardization, increased descriptive data, and authority controlled data elements, in other words towards cataloging data as we know it, are these new metadata schemes just reinventing the wheel? I believe the answer is "no," because they are designed for use with a structural data model that will allow the computer to do what it is good at: computation and connectivity. This, in turn, gives us the potential to develop integrated library systems that can contain multiple databases that use multiple metadata schemes, employ sophisticated relational linkages, and provide unified inter-

```
<?xml version='1.0'?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
      ns#"
        xmlns:dc="http://purl.org/dc/elements/1.0/"
        xmlns:lcnaf="http://loc.gov/naf">

<rdf:Description rdf:about
       ="http://www.asis.org/Bulletin/Oct-97/vellucci.htm">
   <dc:creator>
     <rdf:Description>
       <lcnaf:100>Vellucci, Sherry L.</lcnaf:100>
       <lcnaf:010>nr_87929612</lcnaf:010>
     </rdf:Description>
   </dc:creator>
</rdf:RDF>
```

Figure 3. RDF Example Using DC and LCNAF

faces, all without beginning from scratch. These changes will not happen overnight. They will require catalogers to expand their concepts of authority control, for although the underlying goals will remain the same, the authority control process will change. Catalogers must learn several metadata schemes and organizational structures beyond AACR and the MARC record. They must free themselves from thinking in terms of flat files and linear access and begin to think in terms of multischeme data registries, new record constructs, and relational data models. They must envision a new spectrum of authority control that includes many types of identifiers along with the more familiar names, titles, and subjects. And most critically, catalogers must actively participate in the development of system architectures and data registries. Only this level of activity will ensure that catalogers play a key role in the development of authority control systems for electronic resources. Will metadata ever be totally under control? Probably not, but new frameworks for description and data transfer models based on authority control principles can provide a wider range of data sources for future catalogs.

Works Cited

Borgman, Christine L. 1997. From acting locally to thinking globally: A brief history of library automation. *Library Quarterly* 67: 215–49.

Bourdon, Francoise. 1994. Name authority control in an international context and the role of the national bibliographic agency. *International Cataloguing and Bibliographic Control* 23 (Oct./Dec.): 71–77.

Consortium for the Computer Interchange of Museum Information (CIMI). Accessed July 14, 1999, www.cimi.org.

Day, Michael. 1997. Metadata: Mapping between metadata formats. UKOLN: the UK Office for Library and Information Networking. Accessed July 14, 1999, www.ukoln.ac.uk/metadata/interoperability.

DC-Dot: Dublin core generator. UKOLN: the UK Office for Library and Information Networking. Accessed July 14, 1999, www.ukoln.ac.uk/metadata/dcdot.

Delsey, Tom. 1989. Authority control in an international context. *Cataloging & Classification Quarterly* 9, no. 3: 13–28.

Dublin Core metadata element set: Resource page. Dublin Core Metadata Initiative. Accessed July 14, 1999, http://purl.oclc.org/dc/about/element_set.htm.

Federal Geographic Data Committee. 1998. Content standards for digital geospatial metadata, version 2.0. Accessed July 14, 1999, www.fgdc.gov/metadata/contstan.html.

Gaynor, Edward. 1996. From MARC to Markup: SGML and online library systems. *ALCTS Newsletter* 7, no. 2: A–D. (Insert: From catalog to gateway: briefings from the Cataloging Form and Function Committee, no. 7). Also available at www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html.

Hensen, Steven L. 1989. *Archives, personal papers, and manuscripts.* 2d ed. Chicago: Society of American Archivists.

Hillman, Diane I. 1998. *Dublin Core Metadata Initiative: User guide working draft 1998-07-31.* Accessed July 14, 1999, http://purl.org/dc/documents/wd-guide-current.htm.

International Federation of Library Associations and Institutions (IFLA). Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional requirements for bibliographic records.* München: K.G. Saur. Also available at www.ifla.org/VII/s13/frbr/frbr.pdf.

Jul, Erik. 1997. Cataloging Internet resources: survey and prospectus. *Bulletin of the American Society for Information Science* 24, no. 1: 6–9. Also available at www.asis.org/Bulletin/Oct-97/jul.htm.

Lagoze, Carl. 1996. The Warwick framework: A container architecture for diverse sets of metadata. D-Lib Magazine 2, no. 7/8 (July/Aug). Accessed July 14, 1999, www.dlib.org/dlib/july96/lagoze/07lagoze.html.

Larsgaard, Mary Lynette. 1996. Cataloging planetospatial data in digital form: Old wine, new bottles—new wine, old bottles. In *Geographic information systems and libraries: patrons, maps, and spatial information,* Linda C. Smith and Myke Gluck, ed.. Urbana, Ill.: University of Illinois, Graduate School of Library and Information Science, 17–30.

Library of Congress. Network Development and Marc Standards Office. 1997. The Encoded Archival Description document type definition. Accessed July 14, 1999, http://lcweb.loc.gov/ead.

Meta Tag Builder. c1999. NetPromote. Accessed July 14, 1999, www.metatagbuilder.com.

Miller, Eric. 1998. An introduction to the resource description framework. D-Lib Magazine 4, no. 5 (May). Accessed July 14, 1999, www.dlib.org/dlib/may98/miller/05miller.html.

Murtomaa, Eeva, and Eugenie Greig. 1994. Problems and prospects of linking various single-language and/or multi-language name authority files. *International Cataloguing and Bibliographic Control* 23 (July/Sept.): 55–58.

Phillips, John T. 1995. Metadata: information about electronic records. *ARMA Records Management Quarterly* 29, no. 4: 52–57.

REUSE: A contribution to the enhancement of international bibliographic compatibility. Sponsored by OCLC and the State and University Library of Göttingen, seat of the Regional Library Network for Central and Northern Germany (GBV). Accessed July 14, 1999, www.oclc.org/oclc/cataloging/reuse_project/reuse_final_report.htm.

Shieh, Jackie. 1998. The TEI Header and the Cataloging Rules. In *Final Report, Committee on Cataloging: Description and Access, Task Force on Metadata and the Cataloging Rules.* Chicago: American Library Association. Also available at www.ala.org/alcts/organization/ccs/ccda/tf-tei4.html.

SOLINET, 1999. Community Planet. Accessed July 14, 1999, http://commplanet.solinet.net.

Sperberg-McQueen, C. M., and Lou Burnard. 1994. *Guidelines for electronic text encoding and interchange.* Chicago: Text Encoding Initiative. Also available at http://etext.virginia.edu/TEI.html.

Strawn, Gary L. 1995. LITA/ALCTS Microcomputer Support of Technical Services and Authority Control in the Online Environment Interest Group Joint Meeting, February 5, 1995. *Cataloging & Classification Quarterly* 21, no. 1: 107–9.

Taylor, Arlene G. 1994. The information universe: Will we have chaos or control? *American Libraries* 25: 629–32.

Thibodeau, Sharon G., Daniel Pitti, and Randall K. Barry. 1996. Development of the Encoded Archival Description document type definition. Encoded Archival Description Official Web Site, Library of Congress, Network Development and MARC Standards Office. Accessed July 14, 1999, http://lcweb.loc.gov/ead/eadback.html.

Tillett, Barbara B. 1996. International shared resource records for controlled access. Paper presented at OCLC's Authority Control in the Twenty-first Century: An invitational conference, March 31–April 1, 1996. Accessed July 14, 1999, www.oclc.org/oclc/man/authconf/tillett.htm.

Vellucci, Sherry L. 1998. Bibliographic relationships. In *The principles and future of AACR: Proceedings of the International Conference on the Principles and Future Development of AACR, Toronto, Ontario, Canada, October 23–25, 1997.* Jean Weihs, ed. Ottawa: Canadian Library Association; Chicago: American Library Association, 105–46.

Weibel, Stuart L. 1997. The Dublin Core: A simple content description model for electronic resources. *Bulletin of the American Society for Information Science* 24, no. 1: 9–11. Also available: www.asis.org/Bulletin/Oct-97/weibel.htm.

———. 1999. The state of the Dublin Core Metadata Initiative, April 1999. D-Lib Magazine 5, no. 4 (Apr.). Accessed July 14, 1999, www.dlib.org/dlib/april99/04weibel.html.

World Wide Web Consortium (W3C). 1999. Resource Description Framework (RDF). Accessed July 14, 1999, www.w3.org/RDF/Overview.html.

Younger, Jennifer. 1995. After Cutter: Authority control in the twenty-first century. *Library Resources & Technical Services* 39: 133–41.