4-2007

Application Profile Development for Consortial Digital Libraries: An OhioLINK Case Study

Emily A. Hicks
University of Dayton, ehicks1@udayton.edu

Jody Perkins
Miami University - Oxford

Margaret Maurer
Kent State University - Kent Campus

Notes on Operations

Application Profile Development for Consortial Digital Libraries

An OhioLINK Case Study

By Emily A. Hicks, Jody Perkins, and Margaret Beecher Maurer

In 2002, OhioLINK's consortia of libraries recognized the need to restructure and standardize the metadata used in the OhioLINK Digital Media Center as a step in the development of a general purpose digital object repository. The authors explore the concept of digital object repositories and mechanisms used to develop complex data structures in a cooperative environment, report the findings and recommendations of the OhioLINK Database Management and Standards Committee (DMSC) Metadata Task Force, and identify lessons learned, addressing data structures as well as data content standards. A significant result of the work was the creation of the OhioLINK Digital Media Center (DMC) Metadata Application Profile and the implementation of a core set of metadata elements and Dublin Core Metadata Element Set mappings for use in OhioLINK digital projects. The profile and core set of metadata elements are described.

Introduction

Digital repositories have evolved from relatively simple collections of digital objects with individual metadata schemas to complex online environments needing reliable and flexible metadata structures to accommodate differing demands, platforms, and services. One example of this trend, the OhioLINK Digital Media Center (DMC) developed out of a statewide collaborative environment and continues to be redefined to meet the needs of cooperating libraries.[1] OhioLINK, the Ohio Library and Information Network, is a consortium of eighty-five college and university libraries and the State Library of Ohio. The goal of OhioLINK is to provide easy access to information and swift delivery of materials throughout the state. OhioLINK services include a central online catalog, shared electronic resources, a electronic theses and dissertations center, and an environment for digital project development and access.

By 2002, five years after the DMC was established, the need to restructure and standardize the metadata was clear to OhioLINK staff and member libraries. The DMC provides access to a variety of digital media assets including image, sound, and video files from OhioLINK institutions, other partner organizations, and commercial vendors. A series of subject-specific databases had been created, each with a separate, discipline-appropriate metadata scheme. Little attempt had been made to standardize information across the databases and searching was limited to one database at a time.

OhioLINK's Database Management and Standards Committee (DMSC), composed of technical services representatives from OhioLINK member institutions, appointed the OhioLINK Database Management and Standards Committee (DMSC) Metadata Task Force in spring 2003. The Task Force was charged with providing direction to the DMSC and OhioLINK on the development of the DMC, surveying current and emerg-

Emily A. Hicks (Emily.Hicks@notes .udayton.edu) is Head, Bibliographic Management, and Assistant Professor, Roesch Library, University of Dayton, Dayton, Ohio; **Jody Perkins** (perkintj @muohio.edu) is Metadata Librarian, Miami University Libraries, Oxford, Ohio; **Margaret Beecher Maurer** (mbmaurer @kent.edu) is Head, Catalog and Metadata, Kent State University Libraries and Media Services, Kent, Ohio.

ing metadata standards, and drafting guidelines for the use of metadata in the DMC.

The primary result of the Task Force's work is the OhioLINK DMC Metadata Application Profile.[2] Complex environmental and historical factors and the great diversity of needs within the OhioLINK environment informed the application profile creation process. This paper describes the mechanisms used to foster the evolution of data structures in a cooperative environment and discusses specific decisions and findings that resulted in the creation of an application profile, including the identification of a core set of metadata elements. The paper presents the Task Force's findings, lessons learned, and recommendations, addressing data structures as well as data content standards. Finally, the paper describes the current status of the DMC as well as plans to incorporate the DMC into the OhioLINK Digital Resource Commons (DRC).[3]

A Review of the Literature

Identifying Appropriate Metadata

Several studies have shown that quality metadata is an important component of digital collections. In their article about the challenges of metadata in university digital libraries, Attig, Copeland, and Pelikan assert that successful digital libraries must have a "robust metadata structure that can accommodate and preserve a variety of discipline-specific metadata while supporting consistent access across collections."[4] In a 2004 study of Australian digital collections, Hider finds that respondents think using already established standards when describing digital collections is very important.[5] Bruce and Hillmann point out that while the library community is comfortable with attempting to quantify and measure quality, as evidenced by the acceptance of the BIBCO core

record, this acceptance must take place at the community level, and that "most metadata communities outside of libraries are not yet at the point where they have begun to define, much less measure, quality."[6] Dushay and Hillmann adapt a commercially available visual graphical analysis tool to evaluate metadata, with the aim of developing a tool for efficiently analyzing large databases of metadata.[7]

Broad agreement on what constitutes *good* metadata, or even *appropriate* metadata is difficult. Scalability and relevance have been identified by Intner, Lazinger, and Weihs as features of good metadata as well as "adequate description of the kinds of data elements for which the library's users search."[8] This last factor can vary widely within any consortium's community. Researchers also have found that designing elaborately perfect metadata schemas may not help provide access in the absence of good data. Attig, Copeland, and Pelikan write that they "were forced to ask how little metadata would be required for discovery" and that "this question is particularly important for image data."[9]

According to the National Information Standards Organization (NISO) framework, good metadata is appropriate for the materials in the collection, users of the collection, and intended current and likely use of the digital object; supports interoperability; uses standard controlled vocabularies to reflect the what, where, when, and who of the content; includes a clear statement on the conditions and terms of use for the digital object; is authoritative and verifiable; and supports the long-term management of objects in collections.[10]

Specific guidelines, such as the Computer Interchange of Museum Information's (CIMI) "Guide to Best Practice: Dublin Core" and the Collaborative Digitization Program's (CDP) "Dublin Core Metadata Best Practices," provide a more detailed account of implementing the metadata

component of digital projects.[11] These guidelines typically include element-level guidance on semantics (how to interpret an element), syntax (how to format the data that populates an element), and recommended value domains (what controlled vocabularies, coding schemes, etc. are valid for a given element). The CIMI document guides the implementation of Dublin Core (DC) in a museum environment, presenting element level guidelines for all of the fifteen elements in DC Simple.[12]

Information environments also can heavily affect metadata implementation. Providing access to digital libraries differs significantly from providing access to traditional libraries. Intner, Lazinger, and Weihs note that the very fact that the items being described are online is the "most important and obvious difference."[13] The authors go on to say that:

> Digital libraries are likely to be very large, quickly growing, frequently changing databases; they are likely to be collaborative efforts; they are likely to include more diverse types of materials; and their users do very little searching while they are at the digital library's home institution, if it has only one. As a result, asking a librarian how to find something one believes should be in the database but does not show up in answer to a search query may not be an option. . . . Without standard methods for describing database documents and their contents, maintaining authority control, and so on, access to the documents suffers.[14]

Baca concurs in her article about applying metadata schemas and controlled vocabularies, stating that the metadata standard for cultural heritage institutions must be "appropriate to

the materials in hand and the intended end-users must be selected."[15]

In an article titled "Developing a Metadata Strategy," Agnew details the steps involved in building a metadata repository, including "modeling the information needs of your community, selecting and adapting a metadata standard, documenting your metadata, populating the database and sharing your metadata with other repositories and metadata initiatives."[16] OhioLINK institutions emphasize the importance of that consortial community. Bauer and Carlin explain that the DMC is specifically designed to eliminate barriers to institutional participation and they encourage OhioLINK institutions to focus on "content creation, acquisition and development, thus promoting the true nature of an academic collaborative venture."[17] The impact of this perspective on the quality of the DMC legacy data will be discussed later in this paper.

Cooperative communities have historically struggled to reconcile their independent metadata systems, comprised of legacy data, even in the MARC environment where standards are far more secure. Bruce and Hillmann comment that "legacy data presents special problems for many communities, as it rarely makes a clean transition into new metadata formats."[18] Bishoff and Meagher find no compelling reason to require institutions with legacy data to create new records since "economic reality requires this level of flexibility."[19] Cromwell-Kessler points out that the retrospective conversion of already existing legacy data is "expensive and time-consuming. Where no single standard exists, integration will entail 'translating' from one structured data system to another."[20]

Bishoff and Meagher perceive the challenge for collaborative projects to be the integration of separate collections "using a common set of metadata standards while retaining the unique character of each collection."[21] A 2004 Australian study of digital collections found that almost all of the institutions surveyed valued standardized metadata and federated search functionality and that most were working toward interoperability.[22] Chopey reasons, "Because metadata for digital collections is not likely to be stored for use by any institution except the one creating and maintaining it, the driving force behind the development of metadata standards for digital collections in the future is most likely to be a desire for uniform access methodology across collections."[23] Intner, Lazinger, and Weihs state, "Given the choice between a perfect but unique metadata schema utterly lacking in interoperability and a moderately good schema that gets high marks for interoperability, most experts recommend the latter . . . [because] in a collaborative environment interoperability trumps perfection every time."[24]

If interoperability is the key, how is it attained? Much has been written on the process of cross-walking or data mapping between metadata systems, as well as on the integration of disparate metadata systems within a single database. Cromwell-Kessler says that the process entails "difficult decisions about how to handle complex data issues."[25] Baca writes about the importance of the selection of appropriate metadata schemas and the role of metadata mapping and cross-walks.[26] Bishoff and Meagher discuss how a collaborative project developed a matrix to look at common elements across metadata standards.[27]

The Collaborative Digitization Program (CDP), formerly known as the Colorado Digitization Project, experienced many of these issues.[28] As early as 2000, Allen described the collaborations inherent in the project and the results, noting the great need for good communications and planning within the collaborative environment, stating "[t]he risks relate to quality of the digital objects, digital preservation, and quality of metadata, and these risks must be ameliorated through extensive education and training."[29] The program focuses on the importance of learning through doing, and recognizes that there are unique challenges in cooperative projects.[30] According to Intner, Lazinger, and Weihs, the CDP is currently in the middle of its second strategic plan and doing well.[31]

Attig, Copeland, and Pelikan study the deployment of three separate metadata schema within a single database by creating a merged superset of all the elements in the three standards.[32] Although this exercise proves to be relatively uncomplicated, it does not ensure true interoperability. According to Attig, Copeland, and Pelikan, "The main difficulties concern the *meaning* of the values contained in the elements. . . . They may arise out of contextual differences in the use of language in different disciplines or differences in the role that the data element itself plays in imparting meaning to the values (the hierarchical context). Regardless of the source of the differences, *mapping is about meaning*."[33]

Baca advocates the use of structured vocabularies and thesauri for populating metadata schemas "to increase both precision and recall in end-user retrieval."[34] Metadata created by the contributors can be created more quickly and earlier in the information life cycle for rapidly growing digital collections; the process of metadata creation can more actively involve the contributors in collection development; and the contributors, as experts, can provide more accurate and granular access points.[35] Unfortunately, according to both Chopey and Weibel, this rosy future has not been realized.[36] Weibel calls the prospect of self-archived metadata seductive.[37] Attig, Copeland, and Pelikan contend that, in order to accommodate contributor-created metadata, the requirements for data entry must be kept modest at best.[38]

Few traditional library catalogers have experience outside the MARC and Anglo-American Cataloging Rules paradigm. Data content standards for cultural objects were only recently formalized with the 2006 publication of *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*.[39] Bishoff and Meagher note that one of the major challenges of the CDP is the lack of cataloging expertise, which they consider "a problem for all types and sizes of institutions, not just the small libraries and historical societies."[40] They find that few catalogers participating in the program have experience analyzing and describing digital objects. Chopey observes that the level of granularity within digital collections is often higher than in library catalog.[41]

Caplan's *Metadata Fundamentals for All Librarians* provides an excellent introduction to a variety of metadata schema and serves as a springboard for analysis of available metadata standards.[42] Caplan lays out the principles and practices that underlie most standards and then applies these standards through critical descriptions of various families of metadata schemas. One of the metadata schemas that Caplan describes is Dublin Core (DC). This set of metadata elements was one of the products of an invitational metadata workshop held in Dublin, Ohio, at OCLC, the Online Computer Library Center, in March 1995.[43] The Dublin Core Metadata Initiative's (DCMI) element set has been selected for a multitude of metadata projects, primarily because it supports data mapping and sharing, is Open Archives Initiative (OAI) compliant, and is designed for simplicity of use.[44]

Hider found that most responding libraries used some level of implementation of DC in a 2004 Australian study of digital collections.[45] DC is the metadata element set of choice for the CDP to assure interoperability, although some elements were modified to facilitate the use of DC with digital surrogates of primary source materials.[46] The CDP developed a set of DC-based best practices that provides one example of how to structure an application profile to describe a wide variety of resources in a complex consortial environment.[47]

In a 2004 study of the usage levels of unqualified DC metadata elements in Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) data providers, Ward found that only five of the fifteen elements are used most of the time and that more than half of the eighty-two data providers use only the creator and identifier elements.[48] According to Bruce and Hillmann, "Ward's study indicates that most metadata providers use only a small part of the DC element set, but her study makes no attempt to determine the reliability or usefulness of the information in those few elements."[49] A 2001 survey of DC users by Guinchard indicated that most groups choose DC for its perceived international acceptance, the flexibility of the DC elements, and the probability of future interoperability with other metadata schemes.[50] Critics of the DC element set contend that the fifteen elements are too simplified and calls for expansion have led to the addition of optional qualifiers. Others handle the simplicity issue by including non-DC metadata in addition to DC elements in their projects. In contrast to Ward's study, most of those surveyed by Guinchard use all fifteen DC elements, lending weight to the argument that DC provides a solid foundation for metadata development. The findings support the need for usage guidelines, and some survey participants even call for the development of a DC library application profile.

Baca concludes that there is no "one-size-fits-all metadata scheme" and that therefore the first step is to select the appropriate metadata schema.[51] Cromwell-Kessler notes that metadata systems may be composed of different data elements functioning at different levels, in different ways.[52] Intner, Lazinger, and Weihs suggest that metadata schemas change because new schema develop that have new features, and that standard schema are "nearly always preferred over customized or proprietary schemas that cannot be incorporated easily into a multi-institutional, multi-database, multi-community environment."[53] According to Hider's 2004 survey of Australian digital information providers, the top reasons for choosing a metadata format are:

- most appropriate standard for nature of collection;
- existing standard for non-digital collections;
- community's favored standard;
- government standard;
- interoperability;
- supported by system;
- existing expertise in the standard at the institution;
- requirement for participation in a cross-institution project; and
- simplicity.[54]

Defining Appropriate Metadata Using Application Profiles

Developing application profiles is an important first step in defining appropriate metadata. According to Agnew, "Implementing a core or root schema implies that one's organization will be developing an application profile for the schema. . . . Once one has determined the data elements to be used, the attributes of those data elements, the order in which the data elements will display . . . and whether each element is repeatable, mandatory or optional, it is time to document the application profile."[55] The DCMI Glossary defines an application profile (AP) as:

> a declaration of the metadata terms an organization, information resource, application, or user community uses in its

metadata. In a broader sense, it includes the set of metadata elements, policies, and guidelines defined for a particular application or implementation. The elements may be from one or more element sets, thus allowing a given application to meet its functional requirements by using metadata elements from several element sets including locally defined sets.[56]

Elements can be further refined or narrowed, but not changed. An application profile is not just a model for documentation or for formulating guidelines; it also represents an approach to metadata that is much more flexible and responsive to local needs than is possible when simply adopting someone else's guidelines.

Several reasons to use an application profile are presented by Neuroth and Koch.[57] An application profile provides a standardized way to document the important decisions that have been made about the elements, including content standards and rules for use. Such documentation can facilitate migration, harvesting, and other automated processes. A standard template for documentation makes it easier to maintain consistency across implementations and can assist the development of an overall metadata strategy in the future. An application profile offers a systematic way of developing and sharing a data model. Because an application profile enables tracking across implementations to verify compliance, Heery and Patel suggest it "can provide a basis for different metadata initiatives to work together."[58]

An application profile addresses local needs while still retaining desired levels of interoperability. Dekkers notes that the development of an application profile facilitates the use of multiple schemas because elements can be selected from more than one existing schema or locally cre-

ated and defined.[59] Guidelines unique to a given project or community of practice can be easily documented because, "An application profile is not considered complete without documentation that defines the policies and best practices appropriate to the application."[60] Bruce and Hillmann assert, however, that application profiles are more useful for specialized communities because "[a]pplication profiles, which by their nature are models created by community consensus, demand a level of documentation of practice that is rarely attempted by individual projects or implementers."[61] An application profile provides a framework for a fully developed set of guidelines that contributors can use as a reference or training guide for metadata creators. According to Bruce and Hillmann, "Better documentation at several levels has long been at the top of metadata practitioners' wish list. The first and most general improvement is in the application of standards."[62] Project and collection level application profiles, once archived and made publicly available in an application profile repository, can be used as resources for search terms and other project documentation and by prospective contributors or other project implementers seeking information on projects similar to their own.[63]

Heery and Clayphan note that an application profile, in the form of meta-metadata, also addresses issues of data preservation.[64] In the same manner that technical metadata is required for the ongoing preservation of digital objects, documentation of metadata in the standardized form of an application profile is needed for the preservation of metadata that inevitably will become vulnerable to corruption through the many versions and migrations that have come to be commonplace for digital collections.

Application profiles can be created at different levels of abstraction, ranging from community of practice

guidelines to project level implementations. Three levels are in common use:

- Discipline- or format-based communities of practice seeking to establish a standard set of guidelines specific to a certain discipline or format. Examples include the DCMI, the CanCore Learning Resource Metadata Initiative, and the Video Development Initiative.[65]
- Consortiums or other collaborative groups seeking to establish a common set of guidelines for their members. Examples include the CDP and Canadian Culture Online.[66]
- Local project implementers needing to document local practice, track project specific details, and ensure compliance with other standards. At this level, application profiles are often called data dictionaries and are somewhat different than a full application profile. These local level application profiles include less detail and are more prescriptive since they document all the final choices made for a specific instantiation. Examples include the University of Washington and Miami University.[67]

In "Metadata Principles and Practicalities," Duval et al. support using application profiles to facilitate blending of metadata schemas to accommodate the functional requirements of an application while maintaining a necessary level of interoperability with base schemas.[68] They note, "Metadata modularity is a key organizing principle for environments characterized by vastly diverse sources of content, styles of content management, and approaches to resource description."[69] By combining established metadata schemas and observed best practice, a new application can be developed that meets

local requirements without sacrificing cross-domain interoperability.

Deploying Appropriate Metadata via Institutional Repositories

In 2002, the Association of Research Libraries' Scholarly Publishing and Academic Resources Coalition (SPARC) released "The Case for Institutional Repositories: A SPARC Position Paper," which envisioned an institutional repository (IR) as a "strategic response to systemic problems in the existing scholarly journal system."[70] Lynch defines an IR as a "set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members."[71] Anuradha explains, "Institutional repositories (IR) are digital collections that capture, collect, manage, disseminate, and preserve scholarly work created by the constituent members in individual institutions. They are born out of problems with the current scholarly communication model developed by commercial publishers and vendors."[72] SPARC characterizes these repositories as being institutionally defined, scholarly, cumulative and perpetual, and open and interoperable.[73]

By studying the growth rates in the usage of electronic scholarly information, Odlyzko finds them sufficiently high to predict that "there will be no doubt that print versions will be eclipsed. . . . To stay relevant, scholars, publishers and librarians will have to make even larger efforts to make their material easily accessible."[74] Allard, Mack, and Feltner-Reichert find that "the growth in literature demonstrates that institutional repositories are gaining in momentum throughout academia."[75] In a 2005 study of IR deployment in thirteen nations, Westrienen and Lynch witnessed a great diversity in IRs, and predict that deployment rates will continue to increase.[76] Shearer acknowledges

predicting the long-term success of the IR model is difficult.[77] Chopey notes that successful implementations require broad collaborations of expertise as well as strong guidance from collection curators or compilers.[78] In addition, Lynch observes that the success of IRs depends on institutions recognizing IR as a serious and long-lasting commitment.[79]

Work of the Task Force

The DMSC Task Force's examination of appropriate metadata, application profiles, and institutional repositories revealed challenges for consortial digitization projects such as integrating sometimes disparate collections using common metadata standards, choosing appropriate schemas, and creating good quality metadata. The next steps were to examine the metadata in the DMC, select a base schema, create a set of core metadata elements, and develop an application profile. The remainder of this paper details these decisions, providing recommendations, lessons learned, and conclusions.

Metadata in the Digital Media Center

The DMC was established in 1997 using the Bulldog digital asset management software. When the Task Force began investigating metadata, the DMC contained collections with an eclectic assortment of digital media files of multidisciplinary interest, each with its own unique metadata needs and issues. At the time of this writing, the DMC contains more than 54,000 digital images of art and architecture, more than 1,500 full-length educational videos, and almost 4,000 items in six historic and archival collections. Contributions come from an array of Ohio institutions and arrive in a variety of formats including sound files, digital video, and various standard imaging formats. Commercial

collections—the Encyclopedia of Physics Demonstrations, LANDSAT 7 Satellite Images of Ohio, Sanborn Fire Insurance Maps, Saskia Art History Images, and the ART Collection of art and archaeology objects—are also available through the DMC. Licensing agreements for these databases require OhioLINK to restrict access to individuals associated with an OhioLINK member institution.

Metadata for each collection was supplied by the OhioLINK contributor, a commercial vendor, or harvested by the software. Subject terminologies specific to the genre of the collections, terms used by subject specialists, and terms familiar to patrons desiring access to particular collections of digital media were used. Topical overlap was minimal and the structures and specificity of the terminology varied widely. For example, terms used to describe the photographs in the Wright Brothers Collection were very different from those used to describe the videos in the Encyclopedia of Physics Demonstrations.

The Bulldog software allowed keyword indexing of selected fields within each collection. This indexing was augmented by structured index fields from commercial media products or adapted from the indexing supplied with a project. Descriptive terms for subject searches had to be selected from a pool of terms supplied with the software. The variance in initial metadata and subject terminology resulted in the creation of separate databases, each with metadata appropriate to a specific genre or discipline in addition to the more generic terms supplied by the software.

The limitations of the software ultimately hindered searching of the DMC collections. Content in one collection could not be searched from within another collection, nor could users of the repository expect consistent application of subject terms or consistent search results across the collections. Though a common subject

thesaurus for the DMC was available, it was not apparent from the user interface, nor was the Bulldog thesaurus available to users. By the time the Task Force was formed, a company called Documentum had acquired the Bulldog digital asset management software and was developing software that integrated document management, Web content management, digital asset management, and metadata with functionality to facilitate federated searching and data harvesting. Any new structure would have to address the quality, consistency, and compatibility of the metadata as well as access to the collections. After further examination of the Documentum system, OhioLINK staff decided to look for an open source system that could handle the varied metadata formats, metadata cross-walks, library-specific protocols, and higher education standards needed in today's consortial environment.

Legacy Data in the Digital Media Center

From the beginning, the data structures in the DMC were not apparent or consistent because of the nature of the information. These metadata were created for collections that were designed for different audiences and based on various metadata standards. The need for a cross-disciplinary core set of elements was apparent. Every collection had unique fields and a few common fields that could be mapped to Dublin Core, the Visual Resources Association (VRA) Core, and the Collaborative Digitization Program Core.[80] Multiple types of data structures led to discrepancies between databases and with established standards. For example, the ART Collection data did not follow the standard set by the VRA, and, according to the license agreement, the data had to be mounted as provided. OhioLINK chose to accommodate the needs of a wide variety of contributors rather than risk losing the projects.

While all the databases contained a small number of similar fields, some databases included fields that did not apply to other databases. The Task Force prepared an analysis of metadata in each subject database to determine needs, characteristics, and problems. Initial efforts involved mapping existing DMC metadata and metadata from locally held collections not yet submitted to the DMC into one of several emerging metadata standards. The Task Force then compared the DMC elements to elements used by the Collaborative Digitization Program and Dublin Core. These efforts resulted in "The DMC Core Fields Analysis Document."[81] Further developments of this spreadsheet yielded initial assessments of whether or not each metadata element appeared to be mandatory, required, or optional; whether or not the data field was repeatable; and notations of any issues that appeared to be associated with use of the field.

Selecting a Metadata Standard

Cross-domain interoperability is a common theme throughout digital library research. Digital collections with different architectures, metadata formats, and underlying technologies need common protocols and standards in order to interact. The Task Force agreed that the future of the DMC collections and their growth would depend on finding and adopting a set of metadata standards that would be flexible enough to accommodate the needs of the individual OhioLINK digital collections while facilitating federated searching, a challenge in part because no one had examined the relationships between the DMC databases that would facilitate federated searching. Though procedures (in the format of a proposal form) were in place for submitting collections to the DMC, enforced standards or documentation for establishing new data or metadata structures were not avail-

able to contributors.[82] The Task Force anticipated that a core set of metadata elements accommodating existing and future collections must be developed to facilitate potential development and federated searching. This core set of elements would be anchored in metadata standards and accompanied by a best practices document to assist data compatibility of future DMC collections.

In the preceding few years, there had been an explosion in the growth and development of non-MARC metadata standards. The Task Force considered and rejected a variety of standards for adoption in the DMC. Some standards, such as Encoded Archival Description (EAD) and Metadata Object Description Schema (MODS) were rejected because they were deemed too complicated for non-cataloger contributors.[83] The Text Encoding Initiative (TEI) standard was not considered because of concerns with attaching the metadata directly to the digital object.[84] Several educational standards, including Sharable Content Object Reference Model (SCORM), Learning Object Metadata (LOM) and Metadata for Education Group (MEG), were examined and deemed too specific for this project.[85] The VRA Core Categories also were discussed extensively, but were ultimately discarded as being too cultural object-oriented to accommodate the data.[86] The Task Force ultimately chose an application profile to document the current decisions and to provide the needed framework for a more fully developed set of guidelines in the future.

Selecting a Base Schema

The Task Force needed a base schema that would accommodate the heterogeneous content of the DMC represented by multiple formats, multiple subject areas, and multiple contributors, and simultaneously support federated searching and harvesting. The schema also needed to be interoper-

able with legacy data and be adaptable to change over time. Every effort was made to choose recognized authoritative sources in common use by the digital library community. After a careful review of emerging metadata schemas, best practice documents, and the DMC elements currently in use, the Task Force selected the DC schema as the basis for the core element set because it met the requirements of the DMC environment.

The DC element set was developed with the goals of interoperability, extensibility, and flexibility in mind. Interoperability is important for cross-domain discovery and harvesting. DC provides a high level of interoperability and thus would support federated searching and harvesting. Other standards are too narrow to be applied across all of the DMC collections. The Task Force's work also indicated that all manifestations of existing DMC metadata, as well as selected schemas used in non-DMC collections at OhioLINK member institutions, could be mapped to elements in DC. Dublin Core Simple had been established as an international standard, which increased the possibility that it would come into common use. DC was also the foundation of the OAI-PMH.[87] According to Lagoze, "The OAI approach to metadata harvesting exemplifies the notion of metadata modularization, mandating simple Dublin Core metadata for cross-community interoperability while supporting, in parallel, community-specific metadata for 'drill-down' searching within domains."[88] These trends are important because the larger the community of users for a single standard, the greater the opportunity for resource sharing through harvesting and cross-domain discovery. DC also supports the creation of resource descriptions that are easy to produce and use, which is an important consideration for contributors without access to training or professional catalogers.

Digital Media Center Core Metadata Elements

The Task Force discussed numerous fields as possible core elements and the implications of including and excluding each in the application profile. These discussions were often long and sometimes contentious. Even though most members worked in libraries, a substantial difference of views existed regarding metadata and what steps should be pursued. In the end, the list was narrowed to twenty-two core fields including elements from DC and supplementary elements deemed necessary in the OhioLINK environment. Mapping to the DC element and the DC definition has been retained for those elements drawn directly from the DC element set. Any refinements have been made according to DCMI principles. Table 1 is a list of the core fields and their relationship to the original, the digital manifestation, and OhioLINK asset management. The Task Force viewed these core elements as a starting point for institutions interested in creating metadata for the collections in the DMC. Each institution would have the option to use only the core fields or to include additional fields beyond the core to adequately describe their collections. The creation of subject-related sets of element extensions and additional fields would be possible at any time.

The DMC Core contains six mandatory elements—Title, Creator, Digital Publisher, Asset Type, Object Identifier, and Permissions. Of these six elements, two are system-supplied—Asset Type and Object Identifier—and three are OhioLINK-specific—Asset Type, Object Identifier, and Permissions. By making Title, Creator, and Digital Publisher the only other mandatory elements and by demonstrating that metadata could be as simple or complex as a project warranted, the Task Force hoped to promote widespread adoption of the Core by DMC contributors.

The Title element, defined as a name given to a resource, was the most difficult element to finalize.

Table 1. DMC core elements

Elements related to the original (regardless of format)	Elements related to the digital manifestation	Elements related to OhioLINK asset management
Title*	Digital Publisher*	Collection Name
Creator*	Digital Creation Date	OhioLINK Institution
Contributor	Digitizing Equipment	Asset Type*
Date	Asset Source	OID (Object Identifier)*
Description	Rights	Permissions*
Subject		
Spatial Coverage		
Temporal Coverage		
Language		
Work Type		
Repository Name		
Repository ID		

*Mandatory elements

Source: OhioLINK DMSC Metadata Task Force, "OhioLINK Digital Media Center (DMC) Metadata Application Profile" (May 11, 2004), http://dmc.ohiolink.edu/docs/DMC_AP.pdf (accessed Aug. 11, 2006).

Title

Definition: A name given to a resource. Typically a title will be a name by which the resource is known. It may also be an identifying phrase or object name supplied by the holding institution.

Obligation: Mandatory
Occurrence: Non-Repeatable

Recommended Schemes: none.

Input Guidelines:
1. Identify and enter one Title element per record according to the guidelines that follow.
2. Transcribe title from the resource itself, such as book title, photograph caption, artist's title, object name, etc., using same punctuation that appears on the source.
3. When no title is found on the resource itself, use a title assigned by the holding institution or found in reference sources. If title must be created, make the title as descriptive as possible, avoiding generic terms such as Papers or Annual report. Use punctuation appropriate for English writing.
4. When possible, exclude initial articles from title. Exceptions might include when the article is an essential part of the title or when local practice requires use of initial articles.
5. Capitalize only the first letter of the first word of the title and of any proper names contained within the title.
6. Consult established cataloging rules such as *Anglo-American Cataloguing Rules* (*AACR2*) or *Archives, Personal Papers, and Manuscripts* (*APPM*) for more information.

> **Examples:**
> 1. Channel crew poling ice blocks
> 2. DH4 battle plan and Wright Model C Flyer share air space
> 3. Exhibition flight over Lake Erie
> 4. Great Ballcourt

Maps to DC Element: Title

Figure 1. Title element

Although the Task Force agreed that Title should be mandatory, the occurrence was revised more than once. The Task Force disagreed about whether or not Title should be repeatable or non-repeatable, and whether or not alternate titles should be included in the core elements. If alternate titles were included, should the alternate title be part of the Title element, thus requiring Title to be repeatable, or a separate element? If alternate title was a separate element, should it be a core field? All of these decisions had to be in place before the input guidelines could be finished and the Title element finalized. The Task Force eventually decided to make the Title element non-repeatable and to include any other titles in the additional fields. Additional fields are non-core fields needed for a specific project and are beyond the scope of the application profile document. Figure 1 shows the Title element.

The second mandatory element is Creator, which includes authors, artists, photographers, collectors, or organizations primarily responsible for producing the content of the resource. Entities with a secondary role in the creation process such as editors, illustrators, and preformers are included in the optional Contributor element. Both Creator and Contributor are repeatable fields. Project implementers are instructed to enter names according to established rules (for example, *Anglo-American Cataloguing Rules*, 2nd ed. (*AACR2*), and *Archives, Personal Papers, and Manuscripts*) or use the guidelines outlined in the DMC Metadata Application Profile.[89] The General Input Guidelines state that the same rules or guidelines should be used for names throughout the project profile. The recommended scheme for both elements is the Library of Congress Authorities file.[90]

The Date element contains the creation or modification date or dates of the original resource. Date is required (if applicable) and repeatable. A resource may have several dates associated with the original resource such as creation date, copyright date, revision date, and modification date. The Digital Creation Date element records the date of creation or availability of the digital resource and may be approximated by the agency of creation. This element is required (if available) and non-repeatable. Date maps to DC.date while Digital Creation Date maps to DC.date.available, a refinement of DC.date. The recommended scheme for both elements is ISO 8601, the International Standard for the representation of dates and times.[91]

The Description element is an account of the content of the resource and may include an abstract, table of contents, provenance, or other descriptive text. The Description element holds specialized information that is not included in other elements. Description is required (if available) and repeatable. The Subject element, or topic of the content of the resource, is required (if available) and repeatable. The application profile strongly recommends selecting a value from, or creating values according to, a controlled vocabulary, name authority file, or formal classification scheme to ensure consistency, reduce spelling errors, and improve the quality of search results. Examples include the *Library of Congress Subject Headings* (*LCSH*), Medical Subject Headings (MeSH), and the *Thesaurus for Graphic Materials I: Subject Terms*.[92]

Spatial Coverage describes the location or locations covered by the intellectual content of the resource, not the place of publication. Examples include place names, longitude, and latitude. Recommended schemes for

Spatial Coverage include the Getty Thesaurus of Geographic Names, DCMI Box, DCMI Point, ISO 3166, and *LCSH*.[93] Temporal Coverage refers to the time period covered by the intellectual content of the resource, not the date of publication or digital creation date. The recommended schemes for Temporal Coverage are ISO 8601 and *LCSH*. Both coverage-related elements are optional, repeatable, and map to DC.Coverage, which includes refinements for spatial and temporal coverage.

The Language element records the language of the intellectual content of a resource and is required (if available) and repeatable. Some resources may contain multiple languages while others, such as images, may not contain a language component at all. The recommended scheme for Language is ISO 639-2, a three letter code set for the representation of names of languages.[94]

Work type refers to the manifestation of the original element and is required (if available) and repeatable. The application profile suggests applying terms from an established scheme such as the Art and Architecture Thesaurus or the *Thesaurus for Graphic Materials II: Genre and Physical Characteristics* to ensure consistent usage.[95] Asset Source records the immediate parent or manifestation of the digital object and often will be the same as Work Type. This element is optional and repeatable.

Repository Name lists the organization or institution that holds the original physical object, if applicable. Repository ID holds a number or other identifier for the resource from which the present resource was derived, such as a local accession number. Both of these elements are optional, because some digital resources do not have a repository, and both are repeatable. The Collection Name element records the formal or informal group of objects to which the item belongs. This element is optional and repeatable.

The Digital Publisher is defined as the entity responsible for making the resource available to OhioLINK. Examples include an academic department, corporate body, publishing house, or museum. This element is mandatory and repeatable. If Digital Publisher is the same as Creator or Contributor, the application profile instructs users to enter the information in both elements. This element may or may not be related to the entity listed in the OhioLINK Institution element, which is a consistent reference to the OhioLINK member that contributes the material. OhioLINK Institution is required (if available) and repeatable. Like Creator and Contributor, the recommended scheme for Digital Publisher is the Library of Congress Authorities File.

The Digitizing Equipment element records the equipment or tools used to create the digital object. This element is optional and repeatable. The Rights element records information about rights held in and over the resource. This optional, repeatable field typically contains a rights management statement for the resource or a reference to a service providing the information. Rights information often encompasses Intellectual Property Rights (IPR), copyright, and property rights. The application profile states that if the rights element is absent, no assumptions may be made about any rights held in or over the resource. The Permissions element lists the audience that the publisher agrees to allow access to the content. This mandatory, non-repeatable element has three options—world, state of Ohio, or OhioLINK.

Asset Type records the manifestation of the resource. The software automatically captures this mandatory, non-repeatable element. Values include image, audio, video, or text, and related properties such as file format, file size, and dimensions. This element maps loosely to both DC.Type and DC.Format. Object Iden-

tifier (OID) is a mandatory unique identifier automatically assigned to the digital object that is subsequently used to form a persistent URL.

Creating the OhioLINK Application Profile

Each element in the application profile contains eight different specifications. Four of the specifications are presented in the condensed view of the DMC Core elements in table 2. "Element Name" represents a single characteristic or property of a resource. The "Definition" specifies the type of information required for the named element. In most cases definitions are taken directly from the Dublin Core Element Set. A definition may also contain comments providing additional information or clarification. "Obligation" indicates whether or not a value must be entered. Three types of obligations are used in this application profile. "Mandatory" is defined as a value that must be entered even if it requires the creation of an arbitrary value. "Required (if available)" is defined as a value that must be included if it is available. "Optional" means that it is not necessary to include a value for this element. "Occurrence" indicates whether a single value or multiple values can be included. Two occurrences are used in the DMC Core—repeatable and non-repeatable.

"Recommended Schemes" refers to established lists of terms or classification codes from which a user can select when assigning values to an element. Two types of schemes—vocabulary-encoding schemes, which are controlled lists of words such *LCSH,* and syntax encoding schemes, which indicate that the value must be formatted in accordance with a formal notation, such as how a date is to be entered—may be used. "Input Guidelines" list common conventions and syntax rules used to guide the data-entry process. In the case of system-supplied elements, a brief

Table 2. DMC core elements (condensed view)

Element name	Obligation	Occurrence of values	Mapping
Title	Mandatory	Non-Repeatable	DC.title
Creator	Mandatory	Repeatable	DC.creator
Contributor	Optional	Repeatable	DC.contributor
Date	Required (if available)	Repeatable	DC.date
Description	Required (if available)	Repeatable	DC.description
Subject	Required (if available)	Repeatable	DC.subject
Spatial Coverage	Optional	Repeatable	DC.coverage.spatial
Temporal Coverage	Optional	Repeatable	DC.coverage.temporal
Language	Required (if available)	Repeatable	DC.language
Work Type	Required (if available)	Repeatable	DC.type
Repository Name	Optional	Repeatable	n/a
Repository ID	Optional	Repeatable	DC.source
Digital Publisher	Mandatory	Repeatable	DC.publisher
Digital Creation Date	Required (if available)	Non-repeatable	DC.date.available
Digitizing Equipment	Optional	Repeatable	n/a
Asset Source	Optional	Repeatable	DC.relation.HasFormat
Rights	Optional	Repeatable	DC.rights
Collection Name	Optional	Repeatable	DC.relation DC.relation.IsPartOf
OhioLINK Institution	Required (if available)	Repeatable	n/a
Asset Type	Mandatory (system supplied)	Non-repeatable	DC.format DC.type
OID (Object Identifier)	Mandatory (system supplied)	Non-repeatable	DC.identifier
Permissions	Mandatory	Non-repeatable	n/a

Source*: OhioLINK DMSC Metadata Task Force, "OhioLINK Digital Media Center (DMC) Metadata Application Profile" (May 11, 2004), http://dmc.ohiolink.edu/docs/DMC_AP.pdf (accessed Aug. 11, 2006).

explanation of the process is provided. Two types of input guidelines are provided—general and element-specific. General guidelines that apply to more than one element are located near the beginning of the application profile to cut down on repetition and length of the document. Input guidelines specific to an element are located on the page for that element. "Examples" are provided for each element to illustrate the types of values, conventions, and syntax used for the element. "Maps to DC Element" gives the DC element equivalent, if applicable.

Input guidelines are included to provide a relatively simple way to promote data consistency and assist with data creation while still allowing some flexibility. The application profile was created to accommodate an audience beyond catalogers and others familiar with metadata creation. The Task Force attempted to anticipate questions and to help those unfamiliar with the metadata process plan their projects by providing decision points up front. While anticipating all situations was impossible, every effort was made to assist contributors in metadata creation. External content standards are also referenced as appropriate.

Current Status of the Digital Media Center and the Application Profile

New collections are no longer being added to the DMC and the collections contained in the DMC are being migrated to a new platform called the Digital Resource Commons (DRC), funded by a 2003 Technology Initiatives grant from the Ohio Board of Regents. The OhioLINK DRC is part of the Ohio Commons for Digital Education, a collaborative effort by OhioLINK, the Ohio Learning Network, and the Ohio Supercomputer Center/OARnet to develop digital education resources, services, and capabilities in Ohio. As part of the DRC, OhioLINK is building a general-purpose digital object repository that will accept and share a wider variety of collections and digital objects than the DMC can accommodate. The DRC will be a collection of research and courseware digital repositories connecting to a wide array of existing systems, including Collaborative Learning Environments, portals, and integrated library systems.

All OhioLINK member institutions are entitled to contribute content to the DRC, eliminating "the need for redundant and costly local investments by enabling Ohio colleges and universities to utilize OhioLINK's hardware, software, and staff to create

their own repositories."[96] Individual repositories are customizable, allowing institutions to define how content is contributed and presented. The contributing institutions maintain ownership of the work and control access, allowing rapid dissemination to worldwide audiences or to a single person. The DRC will enhance the quality of education by providing a shared point of access to Ohio's scholarly knowledge. Students will have "a versatile resource for sharing and showcasing . . . research projects as well as accessing course materials, research and learning objects to support their learning."[97] Further collaborations between OhioLINK, Ohio's K–12 community, and other Ohio institutions will enable the DRC to be a foundation of the Ohio education system in the twenty-first century.

The transfer of DMC collections to the new DRC platform is scheduled to be completed by March 2007. The application profile developed by the DMSC Metadata Task Force and described in this paper will continue to be a foundational document for project development in the new system. Contributing members are encouraged to use the application profile during the planning and implementing stages of new projects. The current application profile will be updated in to reflect the DRC environment.

Final Recommendations

Eight recommendations were presented to the OhioLINK Database Management and Standards Committee centering around three broad categories: the need for continued leadership, a call for high-quality metadata development, and the necessity of knowledge sharing. The first recommendation addressed the need for leadership, oversight, coordination, and continuity for DMC metadata. The Task Force recommended that the DMSC develop and document

an overarching metadata strategy to provide a framework for all the metadata related initiatives at OhioLINK. Furthermore, the Task Force recommended that OhioLINK form a body to coordinate metadata-related projects and initiatives, to guide software and tool development, to facilitate metadata harvesting and federated searching, and to keep OhioLINK metadata documentation up-to-date.

The Task Force recognized that the identification of a core set of metadata elements is only a first step. The need for high-quality metadata development will increase in the future. Therefore, the Task Force recommended that OhioLINK develop extended element sets with supporting documentation for various subject and format areas. The Task Force also recommended that OhioLINK develop policies to address legacy data issues to ensure continued usability of older collections.

A group of recommendations addressed issues of training, marketing, and knowledge sharing. The Task Force recommended that OhioLINK host a workshop or conference on metadata and digital collection practices where participants would begin to form a viable OhioLINK metadata practice community. Concurrently, the Task Force recommended the creation of an electronic discussion list for sharing information among this emergent community and current DMC/DRC contributors.

The Task Force proposed the creation of an online, locally developed, wizard-type tool to assist digital collection managers with project planning. After some mildly heated, mostly humorous debate about what to call this tool, the name "MetaBuddy" was chosen. In concept, MetaBuddy is an interactive version of the OhioLINK application profile that could help potential contributors determine the metadata needs of the collection in question. MetaBuddy would lead the project manager through the applica-

tion profile, facilitating the preliminary mapping of existing data structures to the core metadata elements. The collection-specific application profile created in MetaBuddy would then assist OhioLINK programmers with the data mapping of the local collection into the DMC or DRC. The online tool would promote the use of the application profile through its ease of use and adaptability to local needs, promote the use of the DMC or DRC to mount digital collections, and ensure that the standards in the application profile provide consistent, reliable access to OhioLINK's digital collections. The MetaBuddy online tool is currently in development.

The final recommendation addressed the need to expand knowledge of the DMC and DRC throughout the OhioLINK community. The Task Force saw a need to develop and implement a formal marketing strategy to recruit contributors and content and increase end-user awareness and use. The OhioLINK Database Management and Standards Committee is represented on the steering committee of the DRC and the development of the repository is being closely monitored. DMSC members are currently discussing opportunities to increase the awareness and use of the DRC.

Lessons Learned

The Task Force's work was accomplished over twenty months. During that time, a group of people from different institutions and backgrounds collaborated to build a foundation for OhioLINK digital collections metadata. Many lessons were learned. Here are a few of the most significant:

- Standards are still important. Like anything that requires a certain level of compatibility between systems, metadata is standards-driven. Standards provide the foundation for

interoperability. Anyone who wants to increase access to their digital collections—whether through a collaborative project, metadata harvesting, or Google—needs to be aware of a variety of metadata-related standards.

- Standards do not eliminate the need for local decisions. An application profile can help narrow the choices by making recommendations and providing guidelines. However, local decisions will still need to be made for each project.

- It is not necessary to reinvent the wheel with every project. Even though local decisions need to be made for each project, most projects will have common aspects. Find an example of a locally defined application profile or data dictionary for a similar project and adapt it.

- The best and worst thing about metadata is that it does not come with content standards. Traditional MARC is a package deal, complete with a set of standards that are designed to work for everyone. Few people would think of using MARC with a standard other than *AACR2*. The same can not be said about nontraditional metadata. One can pick and choose from a variety of content standards or even create a local variation. This freedom is good when trying to meet locally defined needs; it is bad when aiming for interoperability.

- The metadata universe is large and subject to change. This might be stating the obvious, but keeping this in mind when planning a new project is important. Standards are supposed to provide a certain amount of stability and users may be tempted to become complacent. However, one

should remember that metadata is standards-based, and new standards and technologies are rapidly appearing on the scene that will need to be reconciled with the current standards and technologies. No matter what standards are adopted, being aware of new developments is important. If collections are to be accessible now and in the future, metadata cannot be created in a vacuum.

- Metadata can be as simple or as complex as wanted or needed. Ideally, the need for interoperability, which requires a core of universal elements, is balanced with the needs of a specific collection, project, or community. One way this can be accomplished is through the use of application profiles and extended element sets. However, research shows that few small or independent projects with limited resources have application profiles. Remember that any attempts to standardize the metadata will help with information retrieval and limited access is better than no access.

- Having a cataloging background is useful. The group decision-making process is complex. Catalogers bring certain assumptions to the table about the importance of standards and guidelines that can jumpstart the metadata process, even if they have little knowledge of non-traditional metadata.

- Identifying a set of core elements is an important first step, but it is only the first step. The work accomplished thus far will serve as a foundation for related initiatives within the OhioLINK community. Continued refinement and expansion of this work must continue to meet the changing needs of the consortial community.

Conclusion

After five years of expansion, the OhioLINK DMC metadata needed some standardization to facilitate access to the collections and future growth. Although procedures to submit new collections were in place, no metadata standards or guidelines were available to assist contributors. One of the tenets of the DMC was to eliminate barriers to institutional participation. The legacy of this principle demonstrates one challenge facing consortial repositories. A series of subject-specific databases based on various metadata standards had been created for different audiences. This variety of resources ultimately hindered access to more than one collection at a time. A Task Force was appointed by the OhioLINK Database Management and Standards Committee to investigate metadata schema and best practices documentation. While the Task Force was unable to discover standards and best practices that could be adopted wholesale by OhioLINK for the DMC, the examination of various best practices documentation and standards helped define a core of cross-disciplinary metadata elements. The development of the OhioLINK DMC Metadata Application Profile and subsequent recommendations by the Task Force helped lay a foundation for the creation of quality, consistent, and compatible metadata for future collections contributed to OhioLINK's online repositories. This application profile will help define projects, schemas, and standards for the new OhioLINK DRC to facilitate access for users and training for contributors.

References

1. OhioLINK, The Digital Media Center. http://dmc.ohiolink.edu (accessed July 30, 2006).

2. OhioLINK DMSC Metadata Task Force, "OhioLINK Digital Media

Center (DMC) Metadata Application Profile" (May 11, 2004). http://dmc.ohiolink.edu/docs/DMC_AP.pdf (accessed Aug. 11, 2006).

3. OhioLINK, The Digital Resource Commons, http://drc.ohiolink.edu (accessed Aug. 11, 2006).

4. John Attig, Ann Copeland, and Michael Pelikan, "Context and Meaning: The Challenges of Metadata for a Digital Image Library within the University," *College & Research Libraries* 65, no. 3 (May 2004): 251.

5. Philip Hider, "Australian Digital Collections: Metadata Standards and Interoperability," *Australian Academic & Research Libraries* 35, no. 4 (Dec. 2004). http://alia.org.au/publishing/aarl/35.4/full.text/hider.html (accessed Aug. 11, 2006).

6. Thomas R. Bruce and Diane I. Hillmann, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in *Metadata in Practice,* ed. Diane I. Hillmann and Elaine L. Westbrooke, 238–56 (Chicago: ALA, 2004): 240.

7. Naomi Dushay and Diane I. Hillmann, "Analyzing Metadata for Effective Use and Re-Use," 1–10. http://purl.oclc.org/dc2003/03dushay.pdf (accessed Aug. 11, 2006).

8. Sheila S. Intner, Susan S. Lazinger, and Jean Weihs, *Metadata and Its Impact on Libraries* (Westport, Conn.: Libraries Unlimited, 2006), 189.

9. Attig, Copeland, and Pelikan, "Context and Meaning," 258.

10. NISOFramework Advisory Group, *A Framework of Guidance for Building Good Digital Collections*, 2nd ed. (Bethesda, Md.: National Information Standards Organization, 2004). www.niso.org/framework/Framework2.html (accessed Aug. 11, 2006).

11. Consortium for the Computer Interchange of Museum Information (CIMI), "Guide to Best Practice: Dublin Core." www.cimi.org/old_site/standards/index.html#FIVE (accessed Aug. 14, 2005; site no longer available); CDP Metadata Working Group, "Dublin Core Metadata Best Practices," version 2.1.1 (Denver, Colo.: Collaborative Digitization

Program, 2006). www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf (accessed Feb. 22, 2007).

12. CIMI, "Guide to Best Practice"; Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1: Reference Description" (Dec. 20, 2004). http://dublincore.org/documents/dces (accessed Aug. 16, 2006).

13. Intner, Lazinger, and Weihs, *Metadata and Its Impact on Libraries*, 189.

14. Ibid.

15. Murtha Baca, "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information," *Cataloging & Classification Quarterly* 36, no. 3/4 (2003): 54.

16. Grace Agnew, "Developing a Metadata Strategy," *Cataloging & Classification Quarterly* 36, no. 3/4 (2003): 31.

17. Charly Bauer and Jane A. Carlin, "The Case for Collaboration: The OhioLINK Digital Media Center," in *Digital Images and Art Libraries in the Twenty-First Century*, ed. Susan Wyngaard, 69–86 (Binghamton, N.Y.: Haworth, 2003): 86.

18. Bruce and Hillmann, "The Continuum of Metadata Quality," 241.

19. Liz Bishoff and Elizabeth S. Meagher, "Building Heritage Colorado: The Colorado Digitization Experience," in *Metadata in Practice,* ed. Diane I. Hillmann and Elaine L. Westbrooke, 17–36 (Chicago: ALA, 2004): 35.

20. Willy Cromwell-Kessler, "Crosswalks, Metadata Mapping, and Interoperability: What Does it All Mean?" in *Introduction to Metadata: Pathways to Digital Information*, ed. Murtha Baca, 19–22 (Los Angeles: Getty Information Institute, 1998): 20.

21. Bishoff and Meagher, "Building Heritage Colorado," 19–20.

22. Hider, "Australian Digital Collections."

23. Michael A. Chopey, "Planning and Implementing a Metadata-Driven Digital Repository," in *Metadata: A Cataloger's Primer*, ed. Richard P. Smiraglia, 255–87 (Binghamton, N.Y.: Haworth, 2005): 259.

24. Intner, Lazinger, and Weihs, *Metadata and Its Impact on Libraries*, 189.

25. Cromwell-Kessler, "Crosswalks, Metadata Mapping, and Interoperability," 20.

26. Baca, "Practical Issues."

27. Bishoff and Meagher, "Building Heritage Colorado."

28. Collaborative Digitization Program home page, www.cdpheritage.org (accessed Aug. 5, 2006).

29. Nancy Allen, "Collaboration Through the Colorado Digitization Project," *First Monday* 5, no. 6 (June 2000). www.firstmonday.org/issues/issue5_6/allen/index.html (accessed Dec. 15, 2006).

30. Ibid.

31. Intner, Lazinger, and Weihs, *Metadata and Its Impact on Libraries*.

32. Attig, Copeland, and Pelikan, "Context and Meaning."

33. Ibid., 256.

34. Baca, "Practical Issues," 47.

35. Stuart R. Weibel, "Border Crossings: Reflection on a Decade of Metadata Consensus Building," *D-Lib Magazine* 11, no. 7/8 (July/Aug. 2005). www.dlib.org/dlib/july05/weibel/07weibel.html (accessed Aug. 11, 2006); Suzie Allard, Thura R. Mack, and Melanie Feltner-Reichert, "The Librarian's Role in Institutional Repositories: A Content Analysis of the Literature," *Reference Services Review* 33, no. 3 (2005): 325–36.

36. Chopey, "Planning and Implementing"; Weibel, "Border Crossings."

37. Weibel, "Border Crossings."

38. Attig, Copeland, and Pelikan, "Context and Meaning."

39. Murtha Baca et al., *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (Chicago: ALA, 2006).

40. Bishoff and Meagher, "Building Heritage Colorado," 30.

41. Chopey, "Planning and Implementing."

42. Priscilla Caplan, *Metadata Fundamentals for All Librarians* (Chicago: ALA, 2003).

43. Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set."

44. Agnew, "Developing a Metadata Strategy."

45. Hider, "Australian Digital Collections."

46. Bishoff and Meagher, "Building Heritage Colorado."

47. CDP Metadata Working Group, "Dublin Core Metadata Best Practices."

48. Jewel Ward, "Unqualified Dublin Core Usage in OAI-PMH Data Providers," *OCLC Systems & Services* 20, no. 1 (2004), 40–47.

49. Bruce and Hillmann, "The Continuum of Metadata Quality," 238.

50. Carolyn Guinchard, "Dublin Core Use in Libraries: A Survey," *OCLC Systems & Services* 18, no. 1 (2002): 0–50.

51. Baca, "Practical Issues," 48.

52. Cromwell-Kessler, "Crosswalks, Metadata Mapping, and Interoperability."

53. Intner, Lazinger, and Weihs, *Metadata and Its Impact on Libraries*, 189.

54. Hider, "Australian Digital Collections."

55. Agnew, "Developing a Metadata Strategy," 36, 41.

56. Dublin Core Metadata Initiative, "DCMI Glossary" (Nov. 7, 2005), www.dublincore.org/documents/usageguide/glossary.shtml (accessed July 18, 2006).

57. Heike Neuroth and Traugott Koch, "Metadata Mapping and Application Profiles: Approaches to Providing the Cross-searching of Heterogeneous Resources in the EU Project Renardus," in *DC-2001: Proceedings of the International Conference on Dublin Core Metadata Applications,* ed. Keizo Oyama and Hironobu Gotoda, 122–29 (Tokyo, Japan: National Institute of Informatics, 2001). www.nii.ac.jp/dc2001/proceedings/product/paper-21.pdf (accessed Aug. 18, 2006).

58. Rachel Heery and Manjula Patel, "Application Profiles: Mixing and Matching Metadata Schemas," *Ariadne* 25 (Sept. 2000). www.ariadne.ac.uk/issue25/app-profiles (accessed Aug. 17, 2006).

59. Makx Dekkers, "Application Profiles, or How to Mix and Match Metadata Schemas," *Cultivate Interactive* 3 (Jan. 2001). www.cultivate-int.org/issue3/schemas (accessed Aug. 18, 2006).

60. Dublin Core Metadata Initiative, "DCMI Glossary."

61. Bruce and Hillmann, "The Continuum of Metadata Quality," 253.

62. Ibid.

63. Erik Duval et al., "Metadata Principles and Practicalities," *D-Lib Magazine* 8, no. 4 (April 2002). www.dlib.org/dlib/april02/weibel/04weibel.html (accessed Aug. 17, 2006); Heery and Patel, "Application Profiles."

64. Rachel Heery and Robina Clayphan, "Metadata Application Profiles" (tutorial, "DC-2005: International Conference on Dublin Core Metadata Applications," University Carlos III of Madrid, Spain, Sept. 15, 2005). http://dublincore.org/temp/tutorial5a_eng.pdf (accessed Aug. 18, 2006).

65. Dublin Core Metadata Initiative, "DC-Library Application Profile (DC-Lib)" (Sept. 10, 2004). http://dublincore.org/documents/library-application-profile (accessed Aug. 17, 2006); Norm Friesen, Sue Fisher, and Anthony Roberts, "CanCore Guidelines for the Implementation of Learning Object Metadata," Version 2.0 (Athabasca, Alberta, Canada: Athabasca University, 2004). www.cancore.ca/en/guidelines.html (accessed Aug. 17, 2006); Grace Agnew and Dan Kniesner, eds., "ViDe User's Guide: Dublin Core Application Profile for Digital Video" (Sept. 9, 2001). www.vide.net/workgroups/videoaccess/resources/vide_dc_userguide_20010909.pdf (accessed Aug. 17, 2006).

66. CDP Metadata Working Group, "Dublin Core Metadata Best Practices"; Athabasca University Experts Team, "Canadian Culture Online (CCO) Policy: Metadata Strategy and Development of a Matrix of Metadata Elements" (Athabasca, Alberta, Canada: Athabasca University, 2004). www.canadianheritage.gc.ca/progs/pcce-ccop/reana/sm-ms/metadata_report_e.pdf (accessed Aug. 17, 2006).

67. Metadata Implementation Group, "UW Libraries Dublin Core Data Dictionaries" (Seattle, Wash.: University of Washington Libraries). www.lib.washington.edu/msd/mig/datadicts/default.html (accessed Aug. 17, 2006); Miami University Libraries Digital Library Program, "Frank Snyder Photograph Data Dictionary" (Oxford, Ohio: Miami University Libraries, 2006). http://athena.lib.muohio.edu/wiki/images/d/da/Snyder_Official_Data_Dictionary.pdf (accessed Aug. 17, 2006).

68. Duval et al., "Metadata Principles and Practicalities."

69. Ibid.

70. Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper," *ARL Bimonthly Report*, no. 223 (Aug. 2002). www.arl.org/newsltr/223/instrepo.html (accessed Aug. 17, 2006).

71. Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *Portal* 3, no. 2 (Apr. 2003): 328.

72. K.T. Anuradha, "Design and Development of Institutional Repositories: A Case Study," *International Information & Library Review* 37, no. 3 (Sept. 2005): 169.

73. Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper" [full paper] (Washington, D.C.: Association of Research Libraries, 2002). www.arl.org/sparc/IR/IR_Final_Release_102.pdf (accessed Aug. 18, 2006).

74. Andrew Odlyzko, "The Rapid Evolution of Scholarly Communication" (Ann Arbor, Mich.: PEAK Conference, 2000). www.si.umich.edu/PEAK-2000/odlyzko.pdf (accessed July 30, 2006).

75. Suzie Allard, Thura R. Mack, and Melanie Feltner-Reichert, "The Librarian's Role in Institutional Repositories: A Content Analysis of the Literature," *Reference Services Review,* 33, no. 3 (2005): 333.

76. Gerard van Westrienen and Clifford A. Lynch, "Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005," *D-Lib Magazine* 11, no. 9 (Sept. 2005). www.dlib.org/dlib/september05/westrienen/09westrienen.html (accessed Aug. 17, 2006).

77. M. Kathleen Shearer, "Institutional Repositories: Towards the Identification of Critical Success Factors," *Canadian Journal of Information and Library Science* 27, no. 3 (2002/2003): 89–108.

78. Chopey, "Planning and Implementing."

79. Lynch, "Institutional Repositories."

80. Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set"; Visual Resources Association, Data Standards Committee, "VRA Core Categories," Version 3.0 (Feb. 20, 2002). www.vraweb.org/vracore3 .htm (accessed Aug. 18, 2006); CDP Metadata Working Group, "Dublin Core Metadata Best Practices."

81. OhioLINK DMSC Metadata Task Force, "The DMC Core Fields Analysis Document" (Columbus, Ohio: OhioLINK, 2006). www .personal.kent.edu/~mbmaurer/docu ments/DMCCoreFieldsAnalysisDocu ment.doc (accessed Aug. 17, 2006).

82. OhioLINK, "OhioLINK DMC Proposal Form." http://dmc.ohiolink.edu/ docs/8.03DMCProposalForm.doc (accessed Aug. 18, 2006).

83. Library of Congress, Encoded Archival Description, www.loc.gov/ ead (accessed Aug. 18, 2006); Library of Congress, Metadata Object Description Schema. www.loc.gov/ standards/mods (accessed Aug. 18, 2006).

84. Text Encoding Initiative Consortium, www.tei-c.org (accessed Aug. 18, 2006).

85. Advanced Distributed Learning, "SCORM" (2004), www.adlnet.gov/ Scorm/index.cfm (accessed Aug. 18, 2006); IEEE Learning Technology Standards Committee, "Learning Object Metadata WG12." http:// ieeeltsc.org/wg12LOM (accessed Aug. 18, 2006); UKOLN, "MEG" (2003), www.ukoln.ac.uk/metadata/ education (accessed Aug. 18, 2006).

86. Visual Resources Association, Data Standards Committee, "VRA Core Categories."

87. Open Archives Initiative, "The Open Archives Initiative Protocol for Metadata Harvesting: Protocal Version 2.0 of 2002-06-14" (Oct. 12, 2004). www.openarchives.org/OAI/open archivesprotocol.html (accessed Aug. 16, 2006).

88. Carl Lagoze, "Keeping Dublin Core Simple: Cross-Domain or Resource Description?" *D-Lib Magazine* 7, no. 1 (Jan. 2001). www.dlib.org/dlib/ january01/lagoze/01lagoze.html (accessed Aug. 18, 2006).

89. *Anglo-American Cataloguing Rules,* 2nd ed., 2002 rev. (Ottawa: Canadian Library Assn.; London: Library Assn. Publishing; Chicago: ALA, 2002); Steven L. Hensen, *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries.* 2nd ed. (Chicago: Society of American Archivists, 1989); OhioLINK DMSC Metadata Task Force, "OhioLINK Metadata Application Profile."

90. Library of Congress, Library of Congress Authorities. http://authori ties.loc.gov (accessed Aug. 18, 2006).

91. World Wide Web Consortium, Date and Time Formats (ISO 8601) (Sept. 15, 1997). www.w3.org/TR/NOTE -datetime (accessed Aug. 18, 2006).

92. Library of Congress, Cataloging Policy and Support Office, Library Services, *Library of Congress Subject Headings* 29th ed. (Washington, D.C.: Library of Congress Cataloging Distribution Service, 2006); National Library of Medicine, Medical Subject Headings (MeSH) (July 14, 2006). www.nlm.nih .gov/mesh/meshhome.html (accessed Aug. 21, 2006); Library of Congress, Print and Photographs Division, *Thesaurus for Graphic Materials I: Subject Terms* (TGM I) (1995). http:// lcweb.loc.gov/rr/print/tgm1 (accessed Aug. 21, 2006).

93. J. Paul Getty Trust, Getty Thesaurus of Geographic Names Online (2000). www.getty.edu/research/ conducting_research/vocabularies/ tgn (accessed Aug. 18, 2006); Dublin Core Metadata Initiative, DCMI Box Encoding Scheme (April 10, 2006). http://dublincore.org/documents/ dcmi-box (accessed Aug. 18, 2006); Dublin Core Metadata Initiative, DCMI Point Encoding Scheme (April 10, 2006). http://dublincore.org/ documents/dcmi-point (accessed Aug. 18, 2006); International Organization for Standardization, ISO-3166 Code Lists. www.iso.ch/iso/en/prods -services/iso3166ma/02iso-3166 -code-lists/index.html (accessed Aug. 19, 2006).

94. Library of Congress, ISO 639-2: Codes for the Representation of Names of Languages (June 7, 2006). www.loc.gov/standards/iso639-2/ englangn.html (accessed Aug. 19, 2006).

95. J. Paul Getty Trust, Art & Architecture Thesaurus Online (2000). www.getty .edu/research/conducting_research/ vocabularies/aat (accessed Aug. 19, 2006); Library of Congress, Print and Photographs Division, *Thesaurus for Graphic Materials II: Genre & Physical Characteristics Terms* (TGM II) (2004). http://lcweb.loc.gov/rr/ print/tgm2 (accessed Aug. 19, 2006).

96. OhioLINK, "The Ohio Digital Resource Commons," *OhioLINK Update* 12, no. 1 (April 2006): 1. www .ohiolink.edu/about/update/apr2006 .pdf (accessed Aug. 18, 2006).

97. Ibid., 2.