# Planning and Implementing
# a Metadata-Driven Digital Repository

## Michael A. Chopey

**SUMMARY.** Metadata are used to organize and control a wide range of different types of information object collections, most of which are accessed via the World Wide Web. This chapter presents a brief introduction to the purpose of metadata and how it has developed, and an overview of the steps to be taken and the functional expertise required in planning for and implementing the creation, storage, and use of metadata for resource discovery in a local repository of information objects. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com> © 2005 by The Haworth Press, Inc. All rights reserved.]*

**KEYWORDS.** Information objects, collections, metadata, resource discovery

## METADATA AND RESOURCE DISCOVERY
## ON THE WORLD WIDE WEB

When metadata emerged as a possible solution for resource discovery in a World Wide Web environment in the mid-1990s, many thought that

Michael A. Chopey is Catalog Librarian, University of Hawaii at Manoa Libraries, Hamilton Library, Honolulu, HI 96822 (E-mail: chopey@hawaii.edu).

*255*

an advantage of metadata, as compared to traditional bibliographic data for library catalogs, was that metadata could be created and maintained by individuals with little or no training or experience in cataloging or indexing. (For a concise history of the term and concept "metadata," see Caplan 2003, 1-3.) Indeed the vision behind many of the earliest metadata schemes or initiatives, for example, PICS (Platform for Internet Content Selection (W3C 1997)), or even the use of "META" tags within the HTML standard itself (Weibel 1996), was that resource description could be accomplished by the authors of web documents at the time the documents were created or published on the World Wide Web.

Clearly, this vision has not been realized. In fact, to a large extent, these well meaning early plans for large-scale Internet resource discovery have backfired, in the sense that most commercial Internet search engines will not index the contents of an HTML <META> field in a World Wide Web document because in the universe of documents indexed by commercial search engines, the <META> field is perceived as more likely to contain deliberate misinformation for marketing or promotion purposes than legitimate, useful metadata input by authors to aid in the discovery of their documents' content.

Metadata schemes that were developed with library input similarly began with a focus on simplicity and ease of application. In the development of the Dublin Core Metadata Element Set (DCMES) (DCMI 2003), one of the primary goals has been to make the standard as uncomplicated as possible, without the complex rules for content that are found in the International Standard Bibliographic Description (ISBD) family of cataloging codes (IFLA 2004) and the *Anglo-American Cataloguing Rules, 2nd edition* (*AACR2* 2004), for example, and without the strict rules for required data and data structure that are found in the MARC (MAchine Readable Cataloging) Format for Bibliographic Data (MARC21 2003). In their 1995 report on the OCLC/NCSA Metadata Workshop, which resulted in the establishment of the Dublin Core Metadata Initiative, Weibel et al. (1995) summarized the consensus of the workshop in these words:

> Since the Internet will contain more information than professional abstractors, indexers and catalogers can manage using existing methods and systems, it was agreed that a reasonable alternative way to obtain usable metadata for electronic resources is to give the authors and information providers a means to describe the resources themselves, without having to undergo the extensive training required to create records conforming to established standards.

As one step toward realizing this goal, the major task of the Metadata Workshop was to identify a simple set of elements for describing networked electronic resources. To make this task manageable, it was limited in two ways. First, only those elements necessary for the discovery of the resource were considered. It was believed that resource discovery is the most pressing need that metadata can satisfy, and one that would have to be satisfied regardless of the subject matter or internal complexity of the object.

The assumptions that led to the development of "simple" metadata schemes such as the Dublin Core, and to the belief that they would need to be easily understood and implemented by authors or web publishers without a background in information professions were–and still are–valid. Because of the sheer volume of information published on the World Wide Web, it would be impossible to "catalog" any significant amount of this information with anything but the simplest set of description standards. Library cataloging is time-consuming and labor-intensive and requires mastery of a complex set of standards and rules, but library cataloging is largely a shared enterprise conducted in a much smaller and more manageable universe of information than the World Wide Web. A large percentage of the items that most libraries catalog are collected by hundreds or thousands of other libraries that can use the same bibliographic record, stored in the same format, and acted upon by essentially the same software, so that economies of scale can be achieved as any given item can be cataloged once–by one cataloger at one library–and the bibliographic record for that item can be immediately available to as many libraries around the world as might want to use it in their own catalogs.

No such economies of scale are likely to ever be realized in the information universe of the Internet, of course, because most documents and other information objects available on the Internet are single-instance items, published and made available by a single organization that would ideally be responsible for their maintenance and for providing the metadata that would make the items discoverable.

Furthermore, there have been two distinct but often overlapping efforts in metadata development over the past ten years. One effort has been to develop and set standards to help organize the World Wide Web on a large scale. The other has been to develop metadata schemes particularly suited to the needs of a certain resource description community, such as visual art museums, geospatial data providers, or government

information distributors, or to the characteristics of a certain type of re-
source, such as art images, literary texts, or archival finding aids.

Though the best known result of the 1995 OCLC/NCSA Metadata
Workshop was the Dublin Core Metadata Element Set, a metadata
scheme that has thus far been adopted mostly for the purpose of control-
ling local collections of objects within their own local domains (rather
than for the purpose of making Internet objects discoverable in the
larger context of the Internet), the spirit of that workshop's recommen-
dations were further reaching. The workshop aimed to begin to develop
solutions to the larger problem of meaningful resource discovery on the
Internet by developing a standard for resource description that could
easily be applied across applications on the Internet, and whereby the
resource descriptions created according to that standard could subse-
quently be harvested by "automated tools [that] could discover the de-
scriptions and collect them into searchable databases."

Though the adoption of the Dublin Core on a scale wide enough to re-
alize this vision has not yet occurred–and might not ever occur exactly
in the way originally hoped for–other initiatives (many led by partici-
pants in that very same 1995 workshop) have made great strides in the
past ten years toward achieving interoperability of metadata standards
on the scale of the entire Internet. Local metadata applications are more
and more likely to desire interoperability with other applications, or at
least the ability to link outside themselves to other stores of information,
such as information about a personal name contained in an authority
file, or information about a subject term contained in an external subject
thesaurus.

## *METADATA AND RESOURCE DISCOVERY*
## *IN A LOCAL DIGITAL COLLECTION*

As libraries have demonstrated over the past century or so, it is possi-
ble to develop rules and standards for resource description and access
that can be followed by institutions around the world with collections of
different sizes and scopes, and which can allow for uniformity in the
method of accessing a collection of any size or type. The standards for
resource description and access that have been developed by libraries
should enable a user who has learned how to access a small specialized
collection through its catalog to easily use the catalog of a larger, more
general collection, and vice versa. The desire on the part of libraries for
uniformity in access methodology across collections and the desire to

share bibliographic data for commonly held items have been the two main driving forces behind the development of national and international cataloging standards and formats.

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. As a practical matter, much of the cooperation among institutions that has occurred to date and led to current metadata standards has probably been motivated more by a sense of trepidation about venturing into this entirely new venture alone, and by the tradition of open information sharing among the communities developing metadata standards, than anything else. Planners and implementers of new projects with a metadata component are still considered pioneers, as standards and documentation for many metadata schemes and types of metadata implementations are often still very sketchy as to the specifics of element definitions, data values, and best practices for technical details on such things as storing metadata and designing effective access mechanisms to retrieve it. Fortunately, guidance from metadata pioneers is becoming more available to new planners and implementers. (An example of this, and one that is highly recommended as a source of guidance on the planning and implementing metadata for various types of collections is: Diane I. Hillmann and Elaine L. Westbrooks, eds., *Metadata in Practice* (Chicago: American Library Association, 2004).)

## *NATURE AND TYPES OF METADATA FOR LOCAL DIGITAL COLLECTIONS*

Generally speaking, metadata for discovery of digital information objects within a collection are designed to retrieve a more granular level of resource than that which bibliographic data in a library catalog are designed to retrieve. For example, if a botanical organization published on the World Wide Web a collection of several thousand digital photographs of individual flowers, a library would not likely want to expend the time and effort required to create several thousands of bibliographic records for those individual photographs, nor would library catalog users be well served by having the catalog's subject and title and name indexes cluttered with subject, name, and title entries for thousands of such micro-level resources. The ideal resource discovery scenario from

a general-use library standpoint would be to have one full bibliographic record in the library catalog describing the digital collection, with a URL link from the online catalog to the online botanical collection. The online botanical collection would be organized by metadata that would allow the searcher to conduct searches that are meaningful in that collection, perhaps using subject terminology different or more specialized than would be appropriate in a general-use library catalog.

This has in fact emerged as a common resource discovery scenario, and it is one that seems to work well for information seekers. The challenge in this scenario is that the burden of describing several thousand information objects, newly available for discovery in a manner in which they never had been available before, has fallen not to the expert information organizers in the library, but instead to the botanical organization, which most likely has never attempted to build any kind of an online database of records like the type they now need in order to provide effective online access to the collection. Librarians and information specialists approach this uncharted territory with a firm knowledge of the fundamentals of information organization. They know the access points through which information seekers are likely to approach an information retrieval system, and know how to effectively designate and assign such access points for a collection of a given size and scope. Organizations that wish to provide metadata for the digital objects they produce and collect are likely to have a specialized knowledge of the characteristics and subject content of the objects themselves and the needs of the clientele that seek them, even if their staff does not include experienced information specialists per se.

The most effective and best developed and maintained metadata schemes that have emerged to date are those that combine the expertise of information specialists with the expertise of the producers or curators of the digital information objects and datasets that are sought by Internet researchers in a given discipline or "community." I will discuss the process of metadata creation for digital-object-based collections in which the digital objects themselves, and the metadata that describe and organize them, can be delivered via the World Wide Web to users with standard web browsers, and in which metadata records are created as separate entities from the objects they describe. This model discussed here is a common one for the management of digital image collections, textual document collections, and other collections in which no specialized software (beyond a web browser and accessories typically found in a personal computer) is required to process or display the data that comprise the collection, and in which the Metadata are not an integral part of

the content of the collection for markup and display purposes (as is the case, for example, with Encoded Archival Description (EAD), a metadata scheme used for the markup of archival finding aids). First, a summary of the general considerations for planning and implementing the metadata component for any type of collection, and then on to some more specific considerations for different types of collections.

## PLANNING AND IMPLEMENTING: FUNCTIONAL EXPERTISE REQUIRED

The successful implementation of a metadata-driven digital repository requires the collaboration of staff with expertise in several different areas, including web design and programming, database or retrieval system design, and what can be called "cataloging" for lack of a better term, even though the cataloging function in a metadata context often encompasses more than it would in traditional library or cultural heritage collection cataloging.

In addition to all of this required expertise, in many projects the most important guidance in planning for the creation and use of the metadata will come from the curator(s) or compiler(s) of the collection, whose knowledge of the characteristics of the resources in the collection, and knowledge of how users seek to discover, identify, and select resources from the collection, should inform all of the most important decisions related to the design of the search and browse mechanisms and navigational aids, and the selection of data elements to be included in the metadata records. The cataloging function, at least in the planning stage, cannot be carried out without close consultation with the curators or compilers of the collection, whose guidance is essential in setting the rules for describing the characteristics of a resource and the rules for transcribing data found on, or associated with, the resource. Finally, because many digital collections are described and controlled by metadata on two levels–those that describe the collection's original items in their original formats, and those that describe the digital manifestations of the original items–the expertise of a digital object formatting or digitization specialist is often required to provide the details of the latter. For example, the curator will provide the details of the original item (e.g., art print: sugar lift aquatint; $8 \times 10$ inches), while the digital object formatting specialist (perhaps the staff member overseeing the scanning of the original) will provide the details of the digital surrogate (e.g., file format, file size, resolution).

### Web Design and Programming Function

The web design and programming component is of immediate importance in the planning of any project that uses metadata to facilitate access to an online collection of information objects. Ideally, the curators or compilers of an online collection have already conceptualized the design of the opening or "welcome" screen, the search interface, and navigational aids within the site–and have consulted with the web designer or programmer to ensure that the vision can be realized–even before a metadata scheme has been selected or defined. If the curators or compilers of the collection know that users will want to "enter" the collection by selecting from the opening screen a category of resources to search or browse in, such as "images, 1920-1939," an exhaustive list of such potential categories should be drawn up by the curator and cataloger as a controlled vocabulary list to be included in one of the metadata scheme's defined data elements. Some types of online collections are better accessed by browsing in pre-coordinated indexes or categories, in which case the formulation of controlled vocabularies is a crucial part of the planning and design of the metadata scheme. Other collections are most effectively accessed by post-coordinated searches executed through a search form, in which case a crucial factor in the selection or design of a metadata scheme is the inclusion of all possible data elements a user might wish to search on. As the size of a collection grows, additional data elements–and more specific values within data elements–might become useful for limiting the results of a search or enabling a more specific search, so this fact should be anticipated in defining the data elements to be included in the metadata scheme and the data values to be contained therein. This aspect of the planning is necessarily a collaborative process involving the curators or compilers, whose knowledge of the collection and its users is necessary for anticipating what data elements will need to be defined in the metadata scheme; the cataloger, whose knowledge of metadata schemes and data structuring informs the selection and design of the metadata scheme and the form in which data should be entered to be most effectively retrieved and displayed; and the web designer or programmer, who will design the searching, browsing, and navigational mechanisms according to the curator's and cataloger's specifications.

The web design-programming function is also called upon for the design of the metadata record creation tool that indexers will use to input data and generate metadata records for that project. Though remote-access metadata creation tools are sometimes available via the World

Wide Web for this purpose, many metadata implementations rely on locally defined data elements and data values in their metadata records that cannot be designated in a general-use metadata creation tool. (See for example, the freely available "Dublin Core Metadata Template" provided by the Nordic Metadata Project (available at http://www.lub. lu.se/cgi-bin/nmdc.pl), or the *Connexion®* cataloging tool available to OCLC subscribers at http://connexion.oclc.org/.)

A metadata creation tool is generally set up as a "web form," which might be installed on an indexer's workstation or might be available to indexers via a local or wide area network or the Internet. The indexer uses the form to enter data values appropriate to the object being described in defined fields, and then saves the record, whereupon the record is automatically encoded in the format in which it will be stored, or at least transported to, the online collection's database. The encoding of the metadata might affect how the metadata are retrieved and displayed by the project's search and display mechanisms and how the record is parsed for storage in the collection database. Encoding metadata records in a standard syntax allows a project to exchange records with other metadata applications outside of the project for which the metadata are created. If the indexer is creating a record in Dublin Core, he or she might enter data values into the web form as follows:

| Title | ▼ | St. Francis and the leper: detail of mural by Jose Clemente Orozco in National Preparatory |
| | ▼ | School, Mexico City. |
| | ▼ | |
| Identifier | ▼ | http://128.171.57.100/speccoll/tina%5Fmodotti/jcctm17.gif |
| | ▼ | |
| URI | ▼ | |
| Type | ▼ | black-and-white photographs |
| | ▼ | |
| Local | ▼ | |
| Contributor | ▼ | Modotti, Tina, • 1896-1942. |
| namePersonal | ▼ | |
| | ▼ | |
| Date | ▼ | 1923 |
| created | ▼ | |
| | ▼ | |
| Description | ▼ | RecordType=Work |
| | ▼ | |
| | ▼ | |
| Description | ▼ | Measurements.dimensions=22 x 18 cm. |
| | ▼ | |
| | ▼ | |
| Description | ▼ | Creator.role= Photographer |

when the record is saved, it might automatically be encoded as follows:

```
<?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:dcq="http://purl.org/dc/qualifiers/1.0/">
<rdf:Description about="http://128.171.57.100/speccoll/tina%5Fmodotti/jcctm17.gif">
<dc:title>St. Francis and the leper: detail of mural by Jose Clemente Orozco in National
Preparatory School, Mexico City.</dc:title>
<dc:creator>Modotti, Tina, 1896-1942.</dc:creator>
<dc:date>1923?</dc:date>
<dc:description>recordType=Work</dc:description>
<dc:description>Measurements.dimensions=22 x 18 cm.</dc:description>
<dc:description>Creator.role=Photographer</dc:description>
<dc:description>Location.currentRepository=Charlot Collection, Hamilton Library, University
of Hawaii at Manoa</dc:description>
<dc:description>IDNumber.currentAccession=JCCTM17</dc:description>
<dc:description>Photographer identified by Jean Charlot on verso: Tina
Modotti</dc:description>
<dc:identifier>http://128.171.57.100/speccoll/tina%5Fmodotti/jcctm17.gif</dc:identifier>
<dc:language>und</dc:language>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>namePersonal</dcq:subjectQualifier>
<rdf:value>Orozco, Jose Clemente,1883-1949</rdf:value> </rdf:Description> </dc:subject>
<dc:type>Image</dc:type>
<dc:type>black-and-white photographs</dc:type>
</rdf:Description>
</rdf:RDF>
```

In this illustration, the record is encoded in XML (eXtensible Markup Language) to facilitate parsing of the stored data by the project's storage and retrieval mechanisms, to facilitate display of its elements in a World Wide Web environment if necessary, and to allow for transmission of this metadata to another metadata application if desired. It is also wrapped in RDF (W3C Resource Description Framework) to facilitate the sharing of the metadata record with other RDF-compliant applications and to make the record meaningful within the context of the emerging *Semantic Web* (W3C 2001). The semantics of the metadata in this illustration, e.g., the choice of the word "title" as an element in the metadata scheme, is governed by the Dublin Core Metadata Element Set and also by local rules. The rules governing the form that data values take (e.g., the omission of an initial article in the transcription of the title data, or the use of the controlled vocabulary term "black-and-white photographs" as a data value in the "type.Local" element) are dictated by the project's own local rules for content and transcription, the formulation and documentation of which are part of the cataloging function. The metadata standard might or might not recommend specific vocabulary as the content for any given data element.

The web form used by an indexer to create metadata records might also be programmed to contain pull-down menus from which the indexer selects terms from a controlled vocabulary when inputting content for a given data element. It might also be programmed to contain URL links to an online authority file or thesaurus maintained by an agency outside of the project itself from which the indexer might select authority-controlled names or subject terms to use as data values in certain data element fields.

### *Database or Retrieval System Design Function*

A system that stores metadata and uses it to retrieve information objects associated with the metadata can be designed in many different ways. The design is usually dependent on the nature of the objects being controlled by the metadata, the structure of the metadata itself, and the retrieval and display needs of the system. After XML-encoded metadata are created, its storage in and indexing by the local system might be accomplished in a number of different ways.

As noted, a metadata record can be created by a metadata generation tool that automatically converts indexer-input data into an XML metadata record that can be parsed, stored, and indexed in the local database system, separate from the information object that the record describes. This is one common approach to metadata creation and storage. Another common approach is for the XML-encoded metadata to be embedded in the digital information object itself, usually at the time the digital information object is created.

The former approach allows for more flexibility in the design of a database to store and retrieve the metadata elements. In this approach, what generally happens is that a metadata record is created by a metadata generation tool, and then the record is imported into a local database system. Though the metadata record is most likely encoded in XML for transmission and parsing, in fact when it is imported into a local database system, it is deconstructed and its data elements loaded into the table or tables of a relationship database. The database design and web design-programming functions make certain the data elements stored in these relational database tables are usable for front-end browsing, searching, and navigational tools envisioned by the curators or compilers of the collection.

The latter approach, in which the Metadata are integrated with the digital information object itself, is employed, for example, by systems designed to make use of Encoded Archival Description (EAD

2002) or the Text Encoding Initiative's TEI P4 metadata standard (Sperberg-McQueen and Burnard 2002), two widely known and well documented metadata schemes that use XML not only to encode metadata, but also to encode the text of the document itself. In the case of these "document-centric" systems, as Caplan (2003, 23) calls them, the local system for metadata storage and use can be seen as a "content management system," which integrates all functions pertaining to "authoring, storage and maintenance, query, and presentation." The database or retrieval system design function in this case faces a very different challenge from that faced by the designer of the more freestanding database used to store and retrieve detached metadata. On the other hand, designers of document-centric content management systems are probably more likely to closely follow a pattern established by the metadata standard itself and by other implementers of the standard, and are therefore less likely to have to invent local solutions specific to their own projects and collections.

### *Cataloging Function*

In many ways, the cataloging function is the function most crucial to the success of any metadata application, and there are numerous benefits to having an experienced professional cataloger involved in all stages of the planning and implementation of a local metadata application. While most library catalogers today have probably not studied or been trained in any metadata scheme other than *AACR*2/MARC cataloging, the leap from that expertise to mastering the documentation of any given descriptive metadata scheme should be relatively easy for a cataloger, provided he or she is given sufficient information about the objects being described and the research needs of the collection's users. An experienced cataloger will be able to understand the data element definitions of an existing metadata standard, and, if necessary, define new elements to serve the specific needs of a collection and its users. He or she will be able to determine which characteristics of the objects in the collection need to be described, which data fields should be used to contain the descriptive data and access data for the objects in the collection, and how the data should be input for optimal retrieval, indexing, and display. Working with the curator or compiler of the collection and the web designer, the cataloger can make recommendations for the search, browse, and navigational mechanisms to be made available to users of the online collection, based on what data is available for retrieval in the metadata records and how the data is structured. A cata-

loger can identify an existing subject thesaurus or list of subject terms that is appropriate for the collection, or help build and define a project-specific one. The cataloger should also have a central role in creating any list of controlled-vocabulary terms that might be needed for any given data element in the collection's metadata records, and for defining the terms. Finally, if the collection's metadata will include personal names, corporate body names, place names, or other proper nouns, the cataloger should be involved in designing the reference structure in those indexes when appropriate.

## *CHECKLIST FOR PLANNING AND IMPLEMENTING A LOCAL METADATA APPLICATION*

Table 1 identifies steps in the planning and implementation of a digital repository managed and organized with metadata. Check marks identify the functional expertise required at each stage of the planning and implementation. The areas of functional expertise identified here are: curatorial, digital object formatting, web design and programming, database and retrieval system design, cataloging and indexing, and administrative/managerial. Depending on the size and scope of the collection and the organization sponsoring the collection and retrieval system, some roles might overlap and some might be further subdivided into more specific functions. A committee with representation by at least one staff member from each functional area should be convened at the outset to execute this plan. Decisions made at every stage should be thoroughly documented.

### *Envision the Final Product*

This stage entails planning what the user of the collection will see upon accessing the home page or welcome screen of the collection on the World Wide Web; what search, browse, and navigational options will be available from that screen and others; and, what the user will be directed to upon clicking any given hypertext link (including a hypertext link embodied by a thumbnail image or other graphic, which is a popular means of providing navigation between pages in the WWW environment) on any given page. Although it is best to plan as much as possible of the final product before the construction of the site begins, most esthetic and navigational functions can be adjusted relatively easily at any point later in the development of the repository, so it is not as

TABLE 1

| | Curatorial | Digital object formatting | Web design and Programming | Database/ Retrieval system design | Cataloging and Indexing | Administrative/ Managerial |
|---|---|---|---|---|---|---|
| Envision the final product | ✓ | | ✓ | ✓ | ✓ | |
| Plan appearance and technical characteristics of digital objects | ✓ | ✓ | ✓ | ✓ | | |
| Estimate disk storage requirements | | | ✓ | ✓ | | ✓ |
| Determine staffing requirements | | | | | | ✓ |
| Determine equipment/technical infrastructure required | | | | | | ✓ |
| Project costs | | | | | | ✓ |
| Secure funding | | | | | | ✓ |
| Procure equipment/technical infrastructure | | | | | | ✓ |
| Hire or select staff | | | | | | ✓ |
| Designate metadata scheme and element set | ✓ | ✓ | | | ✓ | |
| Define parameters and input standards for data values | ✓ | ✓ | | | ✓ | |
| Designate exchange syntax for metadata | | | ✓ | ✓ | ✓ | |
| Design and construct metadata creation tool | | | | ✓ | ✓ | |
| Design database for metadata storage and retrieval | | | ✓ | ✓ | ✓ | |
| Assemble rights and credit information | ✓ | ✓ | | | | |
| Train digital object formatting/digitization technicians | | ✓ | | | | |
| Scan/digitize objects; gather already-digitized objects | | ✓ | | | | |
| Design and construct user interface | ✓ | | ✓ | ✓ | | |
| Program search, browse, and navigational functions | ✓ | | ✓ | ✓ | | |
| Inspect site for Web Content Accessibility Guidelines (WCAG) compliance | | ✓ | ✓ | | | |
| Test final product | ✓ | | | | | ✓ |
| Design channels for ongoing feedback from users | | | ✓ | | | ✓ |
| Publicize and release project | ✓ | | | | | ✓ |
| Secure continuing funding | | | | | | ✓ |

crucial to get these exactly right in the planning stages as it is to identify all of the possible access points that users of the repository will desire. Ideally, the person doing the web design under the curator's direction will be the same person who acts as webmaster when the repository is released to the public.

*Esthetics*

Designing what the user will see upon accessing the home page requires a collaboration between the vision of the curator and the skills of the web designer for making that vision a reality. Decisions to be made at this stage include all esthetic considerations from the color of the background, to the size, color, and style of the fonts, to the selection of images, graphics, and special effects to be displayed on the various pages of the online collection. The choices have an impact on the amount of storage space required, and possibly on response time. Most repositories will want to apply at least some level of standard design template to each page the user accesses in the course of navigating and viewing a collection. The standard template might include copyright information, an e-mail link for contacting the webmaster for information or to provide feedback, the name and logo or graphic associated with the repository and/or institution sponsoring it, etc.

*Navigation Functions*

In some digital repositories, the user accesses the entire collection from search and browse functions on the repository's home page. In others, the user first selects from among subject or other categories, and thereby enters a sub-collection to search or browse in. (The "sub-collection" referred to here and elsewhere in this chapter is not necessarily (and for practical purposes, probably rarely is) a "collection" in any real sense. The objects in the sub-collection do not necessarily reside together in any defined space, nor does their metadata. The sub-collection is created dynamically or "on-the-fly" when the user clicks on a hypertext link.) Determining the optimal categories for sub-collections is important mostly for enabling effective browsing, not for enabling effective searching–enabling a *search* of the entire repository should be no more or less difficult to program than a search of some subset of it, even if the metadata for the subsets resides in separate database tables. Furthermore, most repositories will want to offer an option to search the en-

tire collection at any point in a user's journey through it, even when sub-collections have been defined and the user has entered one.

Another important consideration in providing navigational capabilities in a digital repository is to allow a user to move to any part of the collection, or any search screen or index, with a minimal number of mouse clicks from any other part of the collection or from any stage in the retrieval of data. Allowing a user to cache as much of his or her session as possible can also prevent frustration when a user wants to backtrack. Some collections allow users to retrieve the "search history" of a session at any point and either re-execute a search or refine one. Useful feedback from users can be obtained during the pre-release testing period and in an ongoing fashion during the life of the project, but care should be taken during the planning and design stage to anticipate potential navigation problems.

*Browsing Functionality*

In the context of a digital repository, the term "browsing" can refer to the act of viewing the objects in the collection (or surrogates of them, such as thumbnail images of digital photographs, or citations to textual documents) in either a serendipitous fashion or in a manner that is more or less pre-coordinated by the designers of the repository, or, it can refer to the act of browsing an index of access points that is organized in alphabetical or some other order.

Facilitating the former type of browsing is a process that involves the curatorial, web design, and cataloging functions of the project. The curator might map out on paper all of the categories through which a pre-coordinated browse of the collection might proceed. The web designer would enable the hypertext linking between screens and would write the scripts to gather all of the objects in the collection that should be accessible at that level of the browse. These scripts would use terms or phrases contained in the metadata of the target objects to gather all of the relevant objects for a given category. For example, if the curator determined that a useful browse of a collection of objects related to Japanese civilization would begin on the home page with the user selecting a period or era in Japanese history to browse or search within, the curator might define the categories on the home page as: Jōmon culture, Yayoi culture, Tomb Period, Late Yamato, Nara Period, etc. The user would "enter" one of these categories for further browsing or searching by clicking on a text link or icon. That link would contain a script written by the web designer that would execute a search of the metadata in col-

lection database and retrieve every object that pertained to that category. So for example, if the user clicked on the "Yayoi culture" link or icon, a search would be executed that gathered every object in the collection whose associated metadata contained the word "Yayoi" in some predefined data element or elements. The cataloging function would need to know every category that might be defined anywhere in the site, so that the metadata will contain the appropriate terms to gather every object that belongs to that category.

Enabling browsing of indexes, such as name, title, or subject indexes, likewise requires the collaboration of the curatorial, web design, and cataloging functions. The curator should determine the types of indexes to be made available for browsing, and the places within the site where the indexes will be available. The web designer designs the appearance and linking functionality of the indexes, and the cataloger ensures that there are defined data elements in the metadata records that correspond to the indexes selected by the curator.

For example, the curator might determine that useful indexes for a particular textual document collection would be: author name, document title, and subject. The web designer would place these indexes (perhaps in the form of "pull-down menus," for one popular design example) in the pages within the site where the curator has determined that browsing a particular index might be useful, and would write the scripts to populate that index with the appropriate data from a defined field within the metadata, and to execute the gathering of relevant documents when an access point within an index is selected. The access points contained within this type of index come from a defined data element within the metadata records for the collection. For example, an author index might access a data field defined as "Author" or "Creator." The index is dynamic in that every time a new object and its associated metadata are added to the collection, the index expands to include that new access point. It might also be desirable for the index to count the occurrences of a given access point within the index, e.g., the number of documents in the collection written by a given author. An index must also be designed to sort the data that populates it–usually alphabetically, but sometimes chronologically or according to some other order.

The cataloging role is crucial in constructing useful indexes. The cataloger must ensure that the metadata scheme chosen or designed for the collection defines the data element needed for any given index, and that the data element is defined in such a way as to enable browsing (and also other kinds of searching) at the appropriate level of specificity. For example, if the curator determines that indexes and other retrieval

mechanisms should allow a user to search for or browse all names associated with objects in the collection, but also to search for or browse names of persons or corporate bodies performing a particular role, the cataloger needs to define that particular data element accordingly. In the Dublin Core metadata scheme, this might be accomplished by assigning the data element "Creator" with a locally-specified data element refinement, which might be anything from "photographer" to "course architect" to "funding agency." It would then be the responsibility of the web design and database programming functions to ensure that the data is parsed correctly upon import to the collection database, and that search mechanisms and indexes are able to retrieve the data correctly from the database.

In order for an index to be useful, the data that populates it also must be constructed in a consistent manner and one that is logical for sorting. Data consistency is the responsibility of the cataloging function, and is accomplished by creating documentation with clear and thorough rules for data entry, and by effective training of indexers (who actually create the metadata records).

The creation of a subject index can be simple or complex, depending on how large and detailed the list or thesaurus of subject terms defined for the collection is. The most fundamental principle in constructing a subject index is to use a controlled vocabulary. That is, there should only be one authorized subject term or phrase for a given topic or concept. Most controlled vocabulary lists can be made more useful by providing a reference structure to direct users from unauthorized synonyms to authorized forms.

Many general and discipline-specific subject thesauri exist and can be adopted or adapted for use in a metadata-driven digital repository. Examples of general subject thesauri are *Library of Congress Subject Headings* or the *Sears List of Subject Headings*. Discipline-specific thesauri include lists such as the *Art & Architecture Thesaurus* (*AAT* 2000) or the *Thesaurus of Psychological Index Terms* (Gallagher 2004). Adopting an existing thesaurus makes planning easier, but some thesauri are very complicated, and can be applied correctly and effectively only by highly trained and experienced indexers. If the decision is made to construct a local list or thesaurus, the work should be done under the direction of, or in close consultation with, a professional cataloger experienced in the principles of subject analysis and document indexing. There is an international (ISO 1986) standard for the construction of thesauri, but its documentation might be hard to understand and apply

by anyone who is not already very familiar with subject cataloging and document indexing principles.

Once a subject list has been selected or devised, there are still decisions to be made about how subject browsing and searching will be implemented in the collection. A browsable subject index can be constructed in the same manner as a name or title or other index, but since subject terms in a list are likely to have hierarchical and horizontal relationships with other subject terms in the list, the display of results of a subject browse or search is an additional consideration for the curatorial, cataloging, web design, and database/retrieval system design functions. For example, a hypothetical thesaurus for use with a collection of botanical literature might include among its authorized subject vocabulary the term *lichens*. The reference structure for that hypothetical thesaurus might indicate that *lichens* has broader term *Cryptogams* and narrower terms *Ascolichens*, *Epiphytic lichens*, *lichen-forming fungi*, and *rare lichens*, and that *Ascolichens* in turn has narrower terms *Caliciales*, *Graphidales*, *Lecanorales*, *Lichinales*, *Peltigerales*, *Pertusariales*, and *Verrucariales*. If a user elected to browse on the term *lichens*, an effective display of results might be something like the following:

*Broader term:* <u>Cryptogams</u>

**<u>Lichens</u> [22]**

*Narrower terms:*

<u>Ascolichens</u>
<u>Epiphytic lichens</u>
<u>Lichen-forming fungi</u>
<u>Rare lichens</u>

indicating that the collection contains 22 items on the term selected, and pointing the user to related terms that might also be of interest. If the user clicked on one of the other links, a similarly displayed result would appear, placing the selected term within the context of *its* related subject terms. Library catalog software is designed to display subject term relationships in a fashion somewhat similar to this, but the means of accomplishing this with library catalog software involves loading into the library catalog database an entire file of records wherein each term

in the thesaurus has its own record containing all of its broader and related terms. This would be a highly inefficient method for any metadata-driven repository to try to replicate. A more efficient solution for a metadata-driven repository would be to encode the thesaurus in XML and program the display of subject search or browse results to reference the XML document simultaneously as the search or browse queries the database of object-associated metadata.

*Search Functionality*

If good metadata records with well-formed, precisely defined data are created for the objects in a digital collection, and if the metadata records are parsed and stored correctly when they are loaded into a collection database, enabling effective searching of the collection metadata should be a relatively simple matter from a web design standpoint. Inexpensive products are available to web developers for the implementation of web-based search engines that query the contents of a web-based relational database, and the format of these search engines is familiar to most WWW users. So the most important factors in enabling effective and fruitful searching of a metadata-driven digital repository are to identify the attributes of the objects in the collection by which the objects will be sought by users (curatorial function), to define and label each of these attributes precisely as a separate data element in the metadata scheme (cataloging function), and to ensure that the data in each searchable data element field is entered in a manner that ensures effective retrieval (cataloging function). The means of achieving these objectives are discussed further under "Define metadata scheme and element set" below.

### Plan the Appearance and Technical Characteristics of the Digital Objects in the Collection

This step, which requires the collaboration of the curatorial function and the digital object formatting function, entails setting standards for the quality of digital objects in a collection, the technical specifications to which they will be created, and the file formats in which they will be stored. The digital objects in a collection might exist already in digital form before the repository is created, or might be created after the repository has been established. They can be objects that are "born digital" or objects that existed in a different form and were scanned or otherwise converted to digital form for inclusion in the repository. In

setting standards for the technical characteristics of the digital objects in a collection, the curator should consider the following factors, in consultation with the digital object formatting specialist.

### *Quality vs. Size*

The quality (e.g., image resolution in a visual object, or sound clarity in an audio file) of a digital object is often positively correlated with its size. In some types of collections, therefore, the curator must sometimes balance the benefits of creating high quality digital objects with the benefits of creating smaller objects of lesser quality. In setting standards for the size of digital objects in the collection, the curator also takes into account the extra time that it might take a larger item to load on a user's machine, and disk storage requirements on the collection's own servers.

### *Object Renderability*

Ideally, the objects in a digital collection are renderable by a standard web browser without plug-ins or specialized software or accessories. For some types of objects, the need for plug-ins or specialized software or accessories is, of course, unavoidable. In such cases, the curator and digital formatting expert should endeavor to make the objects renderable by the most standard and widely available plug-ins or software possible, and preferably non-proprietary. When other considerations, such as cost and storage, allow, it might be beneficial to users of the collection to offer objects in more than one format.

### *Format Longevity*

An effort should be made to ensure that the format in which objects are encoded or stored are likely to endure. In long-term cost projections, the possibility that the digital objects in the collection might need to be reformatted in the future should be taken into account.

### *Usability of Objects by Users with Disabilities*

Whenever possible, objects should be available in alternative formats so as to be usable by people with disabilities. Information about the requirements of users with different kinds of disabilities can be obtained from the World Wide Web Consortium's Web Accessibility Initiative. Grant funding of digital projects is sometimes contingent upon compli-

ance with government policies for accessibility, such as the ADA (Americans with Disabilities Act) guidelines in the United States.

### Estimate Disk Storage Requirements

Projections of storage requirements should take into account the space required to store the digital objects themselves, the metadata associated with the objects, and the infrastructure of the repository, including search engines, databases, documentation, graphics, etc. They should also allow for data backup.

### Determine Staffing Requirements

The managerial leader of the project should coordinate this activity in consultation with committee members representing each area of functional expertise. Each functional area in the project should be responsible for determining its staffing needs and for writing job descriptions. The creation of a metadata-driven repository can be accomplished in a finite amount of time, or it can be ongoing. Many projects begin with a backlog of existing not-yet-digitized or digitized but not-yet-cataloged materials, and will need more staff for certain functions at startup than in the ongoing phase. In other projects, the size and scope of a collection might grow after it has been implemented, and therefore staffing needs might be greater later.

In web design and database programming, more staff time is likely to be needed in the planning and startup stages than later. In the area of cataloging, the expertise of a professional cataloger will be needed mostly in the planning stages. The cataloger will write documentation and train indexers. In the actual metadata creation stage, however, most of the work in the area of cataloging can be done by indexers. Ideally, all members of the original planning committee will be available after implementation for consultation on maintenance issues.

### Determine Equipment/Technical Infrastructure Needed

For most metadata-driven digital repositories the following equipment and technical infrastructure will be needed:

- Servers, and possibly backup servers, on which to store the digital objects in the collection and the metadata for the objects.

- Computer workstations for indexers, digitization technicians, web design and programming personnel, and database management personnel. These might be workstations dedicated to the work of the project, or workstations they already use for other work they do for the sponsoring institution. Staff using existing workstations might need extra software installed to perform their project functions.
- Network hardware and software to connect workstations to the project's servers.
- Server and database software.
- Digitization hardware such as scanners, digital cameras, analog-to-digital sound recorders, etc.
- Digitization software or other digital object creation software.
- Internet domain name and IP addresses.
- Reference materials, possibly including subscriptions to established thesauri if these are used by indexers.

### Estimate Costs

Cost projections should include staff time, hardware costs, software and licensing costs, telecommunications costs, and the cost of the physical plant where staff will perform their project functions. Each functional area should supply cost estimates for that to the administrative/managerial function.

### Secure Funding

Many digital repositories are funded by grants. The manager should seek the services of an experienced grant writer who is well informed about the mission of the repository and the needs of its users, whether that person is part of the planning team or outside of it. The process of seeking grant funding can refine or even alter the mission or goals of the online collection.

### Procure Equipment/Technical Infrastructure

This stage should be carried out by the manager after funding has been secured. Documentation compiled during the "determine equipment/technical infrastructure needed" should be used.

### Hire or Select Staff

Staff can be selected from within the ranks of the sponsoring institution or from outside of the organization. In either case, a project that se-

cures grant funding can usually apply part of the grant funding toward the salaries of these staff members. Each functional area in the project should have documented its projected staffing needs and written job descriptions in the earlier stage "identify staffing needs." If possible, representatives from each functional area should also be involved in the hiring process. In areas in which extensive training of staff might be required, such as indexing, an effort should be made to hire staff who can accept a long-term commitment to the project, so that re-training can be avoided and optimal efficiency can be achieved from project staff.

### Designate Metadata Scheme and Element Set

Well before this stage is reached (probably during the "envision the final project stage"), the cataloging function of the project will most likely have determined what established metadata scheme will be employed, or, if a locally invented metadata scheme is to be used, the general parameters of that local scheme.

There are many benefits to selecting an established metadata standard, especially if there is any chance that the repository institution might later wish to integrate its collection with other collections, or make its collections searchable through outside applications' search mechanisms. By adopting an established metadata standard, a project may gain access to users' manuals, documentation on recommended best practices for data values and encoding, and possibly tools for data input. A project can also benefit from the experience of other projects using that scheme, as other implementers will often publish usage guidelines and best practices for their own projects in addition to those promulgated by the scheme's maintenance agency. Methods of encoding metadata are often designed with established metadata schemes in mind, so a project might find documentation on methods of encoding metadata more useful and applicable when using an established metadata scheme. The project's web and database programmers are likely to be able to find in electronic discussion forums shareable scripts and other techniques for making use of the project's metadata if an established metadata scheme is employed.

Metadata schemes are sometimes extensible, meaning that a project can adopt the structure and syntax of a metadata scheme, and use some or all of its defined data elements at their broadest definitions, while still allowing for local description needs. In Dublin Core, this is accomplished by means of data element qualifiers (DCMI 2004), which refine the meaning of a data element while still allowing the element itself to

be understood by other applications in its broader, unqualified meaning. For example, the Colorado Digitization Program (CDP), sponsor of the *Heritage Colorado* digital objects repository, has defined in its local metadata scheme (which is an extension of Dublin Core), the qualified element "Format: creation" to express such attributes as file size, compression, or creation software (Bischoff and Meagher 2004). "Format" is a Dublin Core element; "creation" is a CDP-defined local qualifier. In the context of *Heritage Colorado*, the data contained in this field has a meaning that is defined specifically by the element-qualifier combination. If the data were sent to another Dublin Core application outside of *Heritage Colorado*, that application might not be able to "understand" the more specific qualified meaning, but it *would* understand that the data in this field has to do with the associated object's format, and would render it and make it searchable as such. Most digital object repositories will identify some attributes of objects in the collection that should be expressed in the objects' metadata and must be expressed using locally defined metadata extensions.

I have discussed in some detail the functions carried out by *descriptive* metadata–namely, enabling searching, browsing, and navigation within the collection, and expressing information about the attributes and characteristics of an object that a user of the object might need to know–but not discussed were administrative metadata, another important class of metadata for the management of a collection. Administrative metadata express information that is usually of more interest to the repository's staff than to its users, and which might only by viewable or searchable by staff. For example, administrative metadata might include information about the formatting of an object in terms that are more specific and technical than would need to be expressed in the descriptive metadata for the user's purposes, but which would be useful for staff that might need to reformat the object at some point in the future. Other examples of information that might be carried in the administrative metadata for an object are the price the sponsoring institution paid to acquire the object, information about the donor of an object, the name of the indexer who created the metadata record for an object, an accession number for inventory control, etc. The elements to be included in the administrative metadata should be determined by the cataloger in consultation with the curatorial, digital object formatting, and managerial functions.

In this important stage of the planning of a metadata-driven repository, it is crucial that the cataloging function document all decisions made about the metadata scheme to be used in the project, and write

them up in the form of a local metadata standard. This will become the basis for informing much of the database programming and web design of the online repository, and, when expanded with instructions for inputting and encoding the metadata (see below), will become the project's usage guide for metadata creation and manipulation, and probably the most important piece of documentation produced by the project.

### Define Parameters and Input Standards for Data Values

After the metadata scheme and its data elements have been designated and documented, the cataloger should carefully define each element as to the nature of data that should occupy the field labeled with that data element. These definitions should be documented in the usage guide in terms and language that can be understood by the indexers (and possibly digital formatting technicians) who will be creating the project's metadata. The cataloger should also provide detailed instructions for inputting data, specifying the exact form that the data should take (including punctuation, spacing, and capitalization), and guidelines for transcribing data found on the object for which the metadata are being created. For some elements, the cataloger might stipulate that the data value for that element conform to a standard (such as ISO . . . standard for the expression of a "Date" element like "Date.Created") or that the data value be selected from a list, such as a standard list of subject terms or a thesaurus. Some data elements might reference an online registry of some sort, like the "virtual international authority file" envisioned by Barbara Tillett (2001).

### Designate Exchange Syntax for Metadata

The representation of metadata in a syntax that will make the metadata record readily machine-processable is mostly of concern to projects that share metadata with other applications. For these purposes, XML is probably the most widely used exchange syntax, so it is recommended that the metadata creation tool for the type of project discussed here be designed to output metadata records in XML format. Even for a project that does not intend to exchange metadata records with another metadata application, XML can be recommended as a syntax in which to format the project's metadata records because XML is easily parsed by the database programs that will deconstruct the metadata record to store its elements in the repository's database tables.

### Construct Metadata Creation Tool

As shown in the example on p. 263, the project's metadata creation tool will most likely be a web-form based application in which the indexer selects a data element label for each field of data input, and inputs the data for that field in the form specified.

### Assemble Rights and Credit Information

In some types of repositories, it is important that this information be available at the time the digital objects are created, because usage or legal restrictions on some items may require that a watermark about copyright or creator be embedded in the object, or that the downloadability of an item be disabled. Also, in some metadata creation work flows, it might be more efficient to have the administrative metadata input by digital formatting technicians at the time the object is digitized, and rights management and credit information might be a component of the administrative metadata in a project's metadata scheme.

### Train Digital Object Formatting/Digitization Technicians; Scan/Digitize Objects; Gather Already-Digitized Objects

In this stage, technicians will be trained to perform digitization or digital reformatting to the specifications documented digital formatting function in the stage "Plan the appearance and technical characteristics of the digital objects in the collection." Detailed instructions for operating the scanning or other digital formatting hardware should be created at this stage, and used in the training of technicians. Digital formatting technicians might also be trained in the creation of administrative metadata, using documentation created by the cataloging function.

### Train Indexers; Create Metadata

At this stage all of the documentation needed for the training of indexers should have been written by the cataloging function. Documentation can be revised as necessary if it is not easily understood by the indexers, who will most likely not be professional catalogers. The goal should be to allow indexers to create metadata with minimal supervision. In the early stages of metadata creation, all records created by indexers should be stored in a "review file" for inspection by a professional cataloger, who

will file the reviewed record to the live database. As indexers become more experienced, they may be allowed to store records directly to the live database, but the review file should be available throughout the life of the project so that indexers can file records to it when they are uncertain about the data for any given field. Some repositories will use authority-controlled headings (name and/or subject) in their records. In such cases, the indexer should probably be authorized to select already-established headings from the repository's authority file, but not to enter new, unestablished headings. Records requiring headings that have not yet been established should be stored in the review file, from which a professional cataloger will periodically retrieve them and establish the headings in question (and add them to the authority file).

As noted above, metadata record creation can be accomplished efficiently with a web-based form designed by the repository. Indexers should be encouraged to report problems with the form and the record filing process (such as slow response time or bugs), and suggest improvements to make their work easier and the overall metadata creation workflow more efficient. Indexers will be responsible for transcribing data found on the digital objects (or surrogates of them) and also for analyzing the objects in order to record other characteristics and attributes. The workflow should be designed to make it as easy as possible for indexers to ascertain the information they need at the metadata creation stage. For example, if file size is recorded in the metadata for objects in the collection, the file size should be readily ascertainable by the indexer, perhaps provided by the digital formatting function. Many repositories will wish to have the curatorial function provide suggested name headings and subjects and other controlled vocabulary data so that the indexer is not responsible for determining them. Transcription can be problematic in some types of digital collections because, in textual document indexing for example, a data element such as "title" might be ambiguous or difficult to determine. In newspaper indexing, a given article might have several different titles, including a generic or recurring column title, a more specific title, a subtitle, and a running title that changes when the article continues on a different page. A digital book or report might likewise have a title that differs between the title page, the cover, and the running title appearing at the top or bottom of each page in the item. In library cataloging, one title is always chosen as the main title, and the problem of choosing this main title is addressed by a separate set of rules in the cataloging code for every class of material that might be cataloged by a library, and an ordered list within each class for choosing

the "source of information" for the title and other data that might need to be transcribed. A digital repository might adopt this approach in creating its rules for transcription, or the repository might determine that all titles should be transcribed with appropriate refinement of the "title" element, but without any one title chosen as the main title. In any case, the documentation used by the indexers should give clear and unambiguous instructions for identifying every data element that needs to be transcribed (including title, publisher, place of publication or origin, publisher or distributor, date, standard number appearing on the item, etc.), in addition to the instructions for how that data should be transcribed. Some types of digital objects, such as art images or sound files, might have no transcribable data at all. In such cases, the curatorial function will be responsible for supplying much of the data that an indexer will enter in the metadata record. The curatorial function might need guidance from the cataloging function (in the form of documentation created by the cataloging function) on how to determine and supply this data.

### *Import/Attach Metadata*

In some types of collections, the metadata for the digital objects in the collection will be embedded in the objects themselves. In others, the metadata will exist in separately stored metadata records. The model presented in this chapter assumes that metadata will be created in a separate step from the creation or formatting of the digital object, but this does not necessarily mean that the metadata will not be attached to the digital objects after they are created. In either case–whether the metadata are attached to the digital objects or stored separately from them–the workflow shifts from the cataloging function to the database/retrieval system design function at the point at which a completed metadata record is filed to the database. At that point, the database/retrieval system design function might determine (possibly in consultation with the web design function) that it is more efficient to store the metadata records with the objects themselves, or to store them separately. The database/retrieval system design function will also determine how the record is stored in the repository database to ensure efficient retrieval of the data in the records for browsing, searching, and navigational functions of the repository's user interface. Metadata creation will be an ongoing process in many repositories, but a mass of metadata of some size must be created before any of the retrieval functions of the system can be tested.

### *Design and Construct User Interface; Program Search, Browse, and Navigational Functions*

This stage is executed by the web design function according to the vision laid out by the team at the "envision the final product" stage. As the user interface is being constructed, the curatorial function should be monitoring its progress to ensure that the curatorial vision of the final product is being realized. The web design function might present alternative layouts at each stage of the construction for the curatorial function, or perhaps the entire planning and implementation team, to examine and select the best alternative. Many of the search and browse features of the user interface will be designed using web design and database software. This is an advantage not only because it makes the web designer's task easier, but also because it is likely to result in interfaces that are familiar to World Wide Web users, making the repository easier to use and navigate by users. The web designer should be familiar with relevant guidelines and standards for accessibility for people with disabilities, and should program into the user interface alternative means of access wherever they are required.

### *Inspect Site for Web Content Accessibility Guidelines (WCAG) Compliance*

Documentation on the requirements of users with disabilities can be obtained from the World Wide Web Consortium's Web Accessibility Initiative, which maintains links to governmental policies, implementation plans for web accessibility, information on developing organizational policies on web accessibility, information on selecting and using authoring tools for web accessibility, and links to additional resources on web accessibility including resources outside of W3C. Although this documentation will have been consulted by the web design function during the construction of the user interface, a separate step of inspecting the final pre-release product for compliance with Web Content Accessibility Guidelines (WCAG) standards should be undertaken under the direction of the administrative/managerial function. Findings and design changes made at this stage should be documented, and subsequent reviews should be conducted every time a change is made to the user interface that might affect access by users with disabilities.

### Test Final Product

The final pre-release product should be thoroughly tested by the entire planning and implementation team. The curator or manager might also wish to assemble a focus group of target users to test and provide feedback on the product so that changes can be made to the user interface before the product is made available to the public.

### Design Channels for Ongoing Feedback from Users

This stage can be completed at any stage during the planning. It might entail simply attaching a "contact webmaster" link on every page of the site, in which case a member of the post-release repository management team will receive these messages and channel them to the appropriate functional area. Alternatively, a project can decide to allow for more specific feedback at certain places in the user interface, such as a link for reporting data errors, for example, which would be directed to the cataloging function.

### Publicize and Release Project; Secure Continuing Funding

Increasing the visibility of a project among its target user community, and in some cases among the general public, can greatly increase the chances of receiving continuing funding to keep the repository operational and possibly to expand the collection and add affiliate collections to the repository. It can also help the sponsoring institution receive funding for other similar projects it might wish to pursue. Because the design and maintenance of metadata-driven repositories of digital objects is an endeavor that interests many kinds of institutions who do not have any experience in undertaking such a venture, there is usually great interest in presentations by implementers of successful projects at professional organization conferences and workshops. Presenting at such conferences and workshops is one way of increasing the visibility of a project. Another is to publish articles about project planning and implementation in relevant journals and other media.

## WORKS CITED

*Anglo-American Cataloguing Rules, 2nd ed*. (*AACR2*). 2002 revision, 2004 update. Chicago: American Library Association.

Bishoff, Liz and Elizabeth S. Meagher. 2004. Building Heritage Colorado: The Colorado digitization experience. In Diane I. Hillmann and Elaine L. Westbrooks, eds., *Metadata in practice*. Chicago: American Library Association, pp. 22-25.

Caplan, Priscilla. 2003. *Metadata fundamentals for all librarians*. Chicago: American Library Association.

Colorado Digitization Program, *Heritage Colorado*. Available at: http://www.cdpheritage.org/heritage/.

Dublin Core Metadata Initiative (DCMI). 2003. Dublin Core metadata element set, version 1.1: reference description, available at: http://dublincore.org/documents/dces/.

Dublin Core Metadata Initiative (DCMI). 2004. DCMI metadata terms, available at: http://dublincore.org/documents/dcmi-terms/Qualifiers.

Gallagher, Lisa A. ed. 2004. *Thesaurus of psychological index terms*, 10th ed. Washington, DC: American Psychological Association.

Getty Vocabulary Program. 2000. *Art & architecture thesaurus* Los Angeles, Calif.: J. Paul Getty Trust.

Hillmann, Diane I. and Elaine L. Westbrooks, eds. 2004. *Metadata in practice*. Chicago: American Library Association.

International Federation of Library Associations and Institutions (IFLA). 2004. Family of ISBDs, available at: http://www.ifla.org/VI/3/nd1/isbdlist.htm.

International Standards Organization (ISO). 1986. *Documentation: Guidelines for the establishment and development of monolingual thesauri*, 2nd ed., ISO 2788. Geneva: ISO.

Library of Congress. Cataloging Distribution Service. 2004. *Library of Congress subject headings*, 27th ed. Washington, DC: Library of Congress.

Library of Congress. Network Development and MARC Standards Office. 2003. *MARC 21 concise format for bibliographic data*, 2003 concise ed. available at: http://purl.access.gpo.gov/GPO/LPS35317.

Library of Congress. Network Development and MARC Standards Office and Society of American Archivists. 2002. *Encoded archival description (EAD): Official EAD version 2002 website*. Washington, DC: Library of Congress, available at: http://www.loc.gov/ead/.

*Sears List of Subject Headings*. 2004. 18th ed. Bronx, N.Y.: H. W. Wilson.

Sperberg-McQueen, C.M. and Lou Burnard eds. 2002. *Guidelines for text encoding and interchange*. Oxford: Published for the TEI Consortium by the Humanities Computing Unit, University of Oxford.

Tillett, Barbara. 2001. Authority control on the Web. In *Proceedings of the bicentennial conference on bibliographic control for the new millennium: Confronting the challenges of networked resources and the Web, Washington, D.C., November 15-17, 2000*, ed. Ann M. Sandberg-Fox, Washington, DC: Library of Congress, Cataloging Distribution Service, p. 207-220.

United States. Department of Justice. Americans with Disability Act, *ADA homepage*, available at: http://www.usdoj.gov/crt/ada/adahom1.htm.

Weibel, Stuart. 1996. A proposed convention for embedding metadata in HTML. In *W3C workshop on distributed indexing and searching*, May 1996, available at: http://www.w3.org/Search/9605-Indexing-Workshop/reportOutcomes/S6Group2.html.

Weibel, Stuart et al. 1995. OCLC/NCSA metadata workshop report, available at: http://www.oclc.org:5047/oclc/research/conferences/metadata/dublin_core_report.html.

World Wide Web Consortium (W3C). 1997. Platform for Internet content selection (PICS), available at: http://www.w3.org/PICS/.

World Wide Web Consortium (W3C). Resource Description Framework, available at: http://www.w3.org/RDF/.

World Wide Web Consortium (W3C). 2001. Technology and Society Domain. Semantic Web, available at: http://www.w3.org/2001/sw/.

World Wide Web Consortium (W3C). Web Accessibility Initiative homepage, available at: http://www.w3.org/WAI/Resources/.