Developing a Metadata Best Practices Model: The Experience of the Colorado State University Libraries

By:
Patricia J. Rettig
Shu Liu
Nancy Hunter
Allison V. Level

Abstract: The Metadata Best Practices Task Force at the Colorado State University (CSU) Libraries developed a core set of metadata elements and an accompanying data dictionary to facilitate a coordinated metadata management approach for a central digital repository of diverse digital objects. This article describes the rationale for the Task Force and the process used for its work following a look at the background of digitization and past metadata practices at CSU. The article includes a literature review on institutional metadata projects and examples, and it ends with a description of the Task Force's ongoing work and plans for future assessment.

Keywords: Colorado State University Libraries, CSU, data dictionary, digital projects, metadata best practice, Metadata Best Practices Task Force

Journal of Library Metadata, Vol. 8(4) 2008
Available online at http://jlm.haworthpress.com
Article copies available from The Haworth Document Delivery Service:
1-800-HAWORTH.
E-mail address: docdeliver@haworthpress.com
URL: http://www.HaworthPress.com
doi: 10.1080/19386380802656371

Patricia J. Rettig is the Head Archivist for the Water Resources Archive and the Colorado Agricultural Archive in the Archives and Special Collections Department at the Colorado State University Libraries. She earned her MLS from the University of Maryland, College Park, Maryland. (Email: Patricia.Rettig@ColoState.edu).

Shu Liu is the Digital Content Librarian in the Digital Repositories Services Department at the Colorado State university libraries. She earned her MLIS from the University of British Columbia, Vancouver, British Columbia (E-mail: Shu.Liu@ColoState.edu).

Nancy Hunter is a Metadata Librarian in the Metadata and Preservation Services Department at the Colorado State University Libraries. She earned her MLS from the University of Arizona, Tucson, Arizona (E-mail: Nancy.Hunter@ColoState.edu).

Allison V. Level is the College Liaison Librarian for eh Warner College of Natural Resources at the Colorado State university Libraries. She earned her Med in Higher Education Administration and Student Personnel from Kent State University, Kent, Ohio, and her MLS from Emporia State University, Emporia, Kansas (E-mail: Allison.Level@ColoState.edu). The corresponding address for all authors is Morgan Library, Campus Delivery 1019, Colorado State University, Fort Collins, CO 80523-1019.

Developing a Metadata Best Practices Model: The Experience of the Colorado State University Libraries

Introduction

Digitization of local resources is on the rise, and creating metadata for these digital projects is an ever-increasing task in academic libraries. The activity is certainly nothing new at the Colorado State University (CSU) Libraries, which has been engaged in digitization for more than fifteen years. Indeed, because of this long involvement with such projects, contributing librarians grew weary of starting each project from scratch when it came to metadata decisions. The formation of a Metadata Best Practices Task Force in 2007 led to the development of recommendations for local metadata best practices as well as a data dictionary to be used as the basis for the metadata portion of all present and future digitization projects at the CSU Libraries.

This article will discuss the steps the Metadata Best Practices Task Force took in formulating local best practices as well as a data dictionary. Development of such seemingly basic documents does not put the CSU Libraries at the forefront of digital projects; other institutions' best practices documents and data dictionaries were consulted for local needs. However, in the Task Force's research of related issues, no articles were found that laid out steps taken, gave recommendations, or dispensed advice by institutions that had accomplished what the CSU Libraries also sought to accomplish. While some steps described here may be unique to the CSU Libraries, other institutions can learn from the overall process.

The description of the process the Task Force used is preceded by a literature review of metadata-related articles and some background of the CSU Libraries' various digital projects. The article concludes with a look at the next steps the Task Force will take to further implement, evaluate, and maintain the data dictionary in the interest of sustaining a coordinated metadata management approach.

Literature Review

An extensive body of literature on metadata concepts, standards, project applications, and metadata issues has been published. Two recent articles provide a content analysis or survey of literature on metadata articles. Dalton (2004/2005) looks at interdisciplinary projects which used a combination of Dublin Core and other metadata schemas. She provides an annotated bibliography of scholarly articles published between 2000 and 2005 about selected projects. Lopatin (2006) writes a survey of literature on library digitization projects. She includes a section about articles which describe the creation and use of different types of metadata as well as articles which give information about metadata schemes such as the Metadata Object Description Schema (MODS) and the Metadata Encoding and Transmission Standard (METS).

For this article, the literature review will focus on a selected segment of content regarding institutional metadata projects similar to CSU's and related examples in the three categories of project-specific metadata approaches and implementation; metadata assessment, maintenance, and preservation; and metadata sharing.

Project-specific metadata approaches and implementation

There are many articles about metadata approaches, some of which led to consortial best practices and successful implementation of digital projects. One of two such early articles was Bishoff and Meagher (2004) who discussed a statewide digitization experience and the process of building a heritage collection for Colorado. Work on the Colorado Digitization Project began in 1998. Cherry (2004) detailed the development of North Carolina Exploring Cultural Heritage Online (NC ECHO). As an early statewide digitization project, NC ECHO has stressed creation of metadata standards and best practices.

Another early adopter of metadata standardization across projects was OhioLINK's Databases Management and Standards Committee Metadata Task Force, which led to the OhioLINK Digital Media Center Metadata Application Profile. Development of this profile was discussed by Hicks, Perkins, and Maurer (2007). The final recommendations of the task force and lessons learned make this article essential reading.

Guerard and Chandler (2006) extensively documented the Online Archive of California Working Group's procedures and implementation of guidelines and best practices for California cultural collections. Goldsmith and Knudson (2006) described the process of selecting a metadata standard for use in the Los Alamos National Laboratory Research Library's digital object repository, which holds full-text records and complex digital objects. Their article discusses data type handling and compares standards. Lourdi, Papatheodorou, and Nikolaidou (2007) discussed the multi-layer metadata model needed for the variation in materials in their Greek folklore collections at University of Athens.

Ma (2006) provided the metadata implementation steps for digital projects at the Pennsylvania State University Libraries. She also emphasized the need for working together as an organization. Ma wrote:

> Collaboration brings diverse staff expertise together and is the key to metadata implementation. The proposed metadata workflow cannot be realized without the collaboration with all stakeholders. This collaboration is ongoing during the process. Catalogers must learn new standards, certainly, but also must apply much of their longstanding knowledge and many skills in new and interesting ways. Subject librarians must understand both the possibilities for user services presented by these new metadata schemes and the resulting consequences if they are not used. (p. 14)

A coordinated metadata management approach, featuring a matrix organizational structure, is favored, and Ma stated that metadata projects benefit from active and ongoing maintenance.

Metadata assessment, maintenance, and preservation

Issues and models surrounding metadata continue to grow in number and complexity. Foulonneau (2007) discussed information redundancy across digital collections. She stressed the need for detailed and complete information as well as the importance of metadata contextualization. For example, she wrote:

> Similar metadata records may be effectively redundant for many purposes (e.g., retrieval). If records were created for every single page of a digitized book, then the

> difference between two metadata records from this collection might be only a URL (*Identifier* field) and a page number (*Description* field in this case). The metadata properties used are identical, the length of metadata values is similar and metadata values are different. The *Description* property value only differs in a single character (the page number). Our hypothesis was that the presence of such similar records could have an impact on retrieval and selection of records in the context of a digital library system built on top of a large aggregation. (p. 742)

Stvilia and Gasser (2008) included a literature review on information value and cost models of metadata. They also discussed value-based metadata quality assessment. They suggested that creating metadata and establishing models and best practices are not enough; added to that is a need for further maintenance. LeBlanc and Kurth (2008) outlined workflow models for ongoing metadata maintenance.

A key step in metadata development was the creation in 2003 of the PREMIS (PREservation Metadata: Implementation Strategies) Working Group, an international working group charged with developing a core set of metadata elements for digital preservation. Caplan and Guenther (2005) described the evolution of the group and the continuing work toward developing a final report.

Metadata sharing

One of the benefits to online digital collections is worldwide access to their information and content. Creating quality metadata can be resource intensive but worth the investment if it is easily available for sharing or harvesting. Sharable metadata should be useful and valuable outside the local context. The Open Archives Initiative (OAI) Protocol provides transferability for metadata. Medeiros (2006), Shreeves, Riley, and Milewicz (2006), and Elings and Waibel (2007) all described the importance of metadata sharing and discussed issues with metadata aggregation. Simeoni, Yakici, Neely, and Crestani (2008) focused on a content-based distributed information retrieval approach which, they stated, "occupies middle ground between content crawling and distributed retrieval" (p. 12).

As the literature mentioned above outlines, metadata is a critical component in digital projects, and many universities and organizations are implementing digital projects with metadata schema. Like many of these other institutions, the CSU Metadata Best Practices Task Force aimed to implement a coordinated metadata management approach for a central digital repository.

Background

Colorado State University is a land-grant institution with a current enrollment near 25,000. The CSU Libraries is an Association of Research Libraries institution with more than 2 million books, more than 30,000 electronic resources, and just over 100 faculty and staff. Digital projects at the Libraries have evolved for more than fifteen years, during which nine collections of nearly 4,000 digital objects have been created and made accessible online.[1] Metadata has been an essential component of the projects from the beginning for the following reasons:

- Metadata is indispensable in terms of enhancing search accuracy, aiding evaluation of a particular resource, and facilitating search engine harvesting of digital resources;
- For non-textual materials such as images, metadata is essential for retrieval; and
- Many digital asset management systems allow for fielded searching, which would not be possible without structured metadata.

Reviewing two projects key to the evolution of metadata application at the CSU Libraries will help illustrate the arrival at the formation of the Metadata Best Practices Task Force. The International Poster Collection was the first digital project at the CSU Libraries and remains ongoing. The project began in 1991 as a collaboration between the Libraries and the CSU Department of Art, which holds the Colorado International Invitational Poster Exhibition every two years. One copy of each exhibited poster becomes part of the Libraries' collection, an addition of around 200 posters every other year. Cataloging of the posters began soon after the first acquisition, with fields determined by a library cataloger in conjunction with graphic artists in the Art Department. Library of Congress name and subject authorities were used in the cataloging records, which were developed according to local needs, not national standards. Some local controlled vocabulary lists were created as well for descriptors particularly of interest to graphic design artists. Eventually, the cataloging records were posted on the Libraries' Gopher site. As the Internet evolved and websites became the norm, the poster records and thumbnail images of post-1993 posters were posted as hundreds of static web pages.

In the summer of 2000, a project cataloger was assigned to upgrade the poster collection. Images were moved to a digital content management system, and cataloging records were migrated into Dublin Core (DC), which was barely two years old. Since DC was designed for simplicity of use and flexibility of application, it was deemed appropriate. Most fields of the qualified DC were easily applied, and some customized local fields were utilized as well. At this time, in addition to descriptive metadata, fields for administrative metadata were also established. As the standards for that were just emerging, the draft NISO standard "Technical Metadata for Digital Still Images" was applied.[2]

Concurrent with the migration of the poster collection to a database, a second digital project was requiring the same practices. The Warren and Genevieve Garst Photographic Collection was donated to the Libraries in fall 1999. In December of that year, the Libraries received a $5,000 grant to digitize and make available 1,000 of the 20,000 slides of animals from around the world. Students majoring in zoology were hired to select images and create metadata. They worked with cataloging librarians to establish Dublin Core fields and ensure compliance with the grant requirements. The Libraries' first employee with the title "metadata librarian" was hired in February 2001 and provided guidance for the project's metadata. Because of the collection's subject nature, a number of local fields were created to provide a number of biological fields about the animals, such as Genus Species, Common Name, Geographic Range, Habitat, and Diet.[3]

With these two digital projects underway and more on the horizon, the Libraries investigated content management software and purchased CONTENTdm (CDM) in mid-2001.[4] The new International Poster Collection website debuted with CDM as its engine in September, and the Garst website went live in October. Over the next six years, six small digital projects in the range

of a few dozen items (some compound objects) were completed and loaded into CDM. These included the Germans from Russia Collection, the Colorado's Waters Digital Archive, the Celebrate Undergraduate Research & Creativity Digital Showcase, the AgNIC Carnations & the Floriculture Industry Collection, and the Dot Carpenter Virtual Exhibit.[5] Because the software allowed digital collections to be siloed, metadata fields could be easily customized for each collection. For better or for worse, this allowed each project to determine its own metadata needs independently.

The diversity of subjects and material types involved in these digital projects—as well as the fact that they were discrete projects—were just some of the challenges to the Libraries' application of metadata. Staffing varied from project to project, as did funding sources, with some grants having their own metadata requirements. Additionally, projects were started for different reasons, with different goals. Fundamental differences in project goals affect metadata decisions, as do projects' intended audiences. University students and faculty are the primary targets for the Libraries' digital projects, but broader audiences are as diverse as K-12 students, graphic design artists, water resources professionals, and genealogists. Creating custom metadata to a greater or lesser extent for these diverse groups has challenged digital project managers. Consequently, each project typically started from scratch to determine what metadata was needed.

With the many factors challenging the CSU Libraries' application of metadata over the years, divergent solutions led to a variety of metadata practices. When, in late 2006, the Libraries purchased DigiTool, a digital assets management system, to facilitate the implementation of a central digital repository on campus, a new challenge was introduced.[6] The new software provided more flexibility, but it took a less siloed approach to collections, so metadata had to be more uniform across all digital objects. Also, the Libraries decided to migrate all previous digital collections from CONTENTdm to DigiTool, thus affecting locally customized fields. This, along with twelve digital initiatives planned for the Libraries in fiscal year 2007-08, made clear the need for a coordinated metadata management approach.

Task Force Formation

In spring 2007, the metadata librarian in Metadata and Preservation Services and the digital content librarian in Digital Repositories Services independently approached the coordinators of their respective departments seeking permission to develop a core set of metadata elements for digital projects and the developing institutional repository. After the department coordinators brought this request forward, the Assistant Dean for Digital Services drafted a charge for a Metadata Best Practices Task Force. She assigned the metadata librarian, the digital content librarian, and the head archivist in Archives and Special Collections to the Task Force. A college liaison librarian in the College Liaisons Department expressed interest and was added to the Task Force shortly thereafter.

All four members brought experience building metadata to the Task Force: the first three had developed metadata for most of the digital collections at the Libraries, and the college liaison had experience building the Colorado Agriculture and Rural Life Bibliography database. She also contributed metadata for the Colorado Rangelands portion of the Rangelands West database. Additionally, each brought unique experiences and perspectives to the Task Force. The metadata

librarian was a cataloger as well as a member of the Collaborative Digitization Program's Metadata Working Group. The digital content librarian, in addition to her background as a metadata librarian, was a member of the team developing the new CSU Digital Repository and brought a deeper understanding of DigiTool and its metadata requirements. The archivist brought cataloging experience as well, but her unique contribution was her familiarity with the archival materials involved in many of the digitization projects and their special needs. The liaison librarian provided the important perspective of the end user and the librarians who assist them. This diverse combination proved valuable in the Task Force's deliberations.

At the beginning of the Task Force's work, the Libraries had three digital initiative priorities: a joint-pilot of electronic theses and dissertations (ETD) with the CSU Graduate School, a collection of Atmospheric Science Papers published by CSU faculty members, and an archival collection of historic university photographs being digitized in-house. Plans were to deposit these materials in the new CSU Digital Repository as showcase collections. Processing these materials within a short timeframe generated a pressing demand for a local metadata standard to minimize metadata planning efforts and maintain metadata consistency within the repository. Thus, these priorities influenced the work of the Task Force.

Task Force Work

In June 2007, the Assistant Dean for Digital Services gave the Metadata Best Practices Task Force the following goals:

> 1) To identify metadata standards and schemes currently in use for local digital projects and determine how these might integrate for the future;
> 2) To recommend the implementation of best practices for metadata for digital projects, including consideration of core vs. specific metadata approaches, approaches that take advantage of existing metadata when available, and social networking technology; and
> 3) To provide ongoing support for compliance with best practices that employ technology and the community of individuals who are involved in digital projects.[7]

These goals addressed the request by the metadata and digital content librarians to ease the metadata planning process and provide a baseline from which to build metadata. By having one core set of metadata, the Digital Repositories staff could build the DigiTool interface-- including which fields to display in various views--with a consistent look and feel regardless of the digital collection. While some modifications may be necessary for particular projects, the highly customized metadata such as was previously created would be a thing of the past. Additionally, the investment in time to design metadata for a project would be minimized.

To accomplish these goals, the Task Force was charged with specific tasks divided into the following phases to address primary concerns of the CSU Libraries Strategic Plan:

I) To research and identify a set of core metadata elements for local use;
II) To develop a data dictionary to define local metadata best practice;
III) To conduct an inventory of local digital resources and identify how they are currently managed and accessed;

IV) To investigate and provide a report detailing local human and technology resources necessary to comply with recommended metadata best practice; and

V) To provide ongoing support for library metadata issues.

It may have been more logical to begin with Phase III; however, due to the pressing need for a local metadata standard and a short timeline, the Task Force started with Phases I and II instead. Because three of the four members had been metadata leads for past digital projects, they already possessed knowledge of current local metadata standards and schemes. For this article, Phase I and Phase II will be described in detail.

Phase I

Activities in Phase I are described by specific tasks accomplished within this phase.

1) Research metadata standards and schemes applicable to all digital projects and unique to specific disciplines, and identify technical metadata requirements for multi-media digital objects.

The Task Force identified and researched the following metadata standards and schemes applicable to the local environment:

- *Dublin Core Metadata Element Set (version 1.1)* and *Dublin Core Qualifiers* (Dublin Core Metadata Initiative [DCMI], 2008 and 2005, respectively)
- *Dublin Core Metadata Best Practices (version 2.1.1)* (Collaborative Digitization Program, 2006)
- *ETD-MS: An Interoperability Metadata Standard for Electronic Theses and Dissertations (version 1.00, revision 2)* (Networked Digital Library of Theses and Dissertations, 2001)
- *VRA Core 4.0* (Visual Resources Association, 2007)
- *IEEE 1484.12.1-2002- Draft Standard for Learning Object Metadata* (Institute of Electrical and Electronics Engineers, 2002)
- *ANSI/NISO Z39.87-2006 - Data Dictionary: Technical Metadata for Digital Still Images* (National Information Standards Organization, 2006)
- *Data Dictionary for Preservation Metadata (version 1.0)* (PREMIS Working Group, 2005)

The Dublin Core Metadata Initiative has been in existence since 1995 and is broadly used for describing and facilitating searching of web resources. As a “lowest common denominator” standard, Dublin Core (DC) is flexible and easy to apply. It is particularly useful in addressing cross-collection and cross-discipline searching. DC and the Collaborative Digitization Program (CDP) DC Metadata Best Practices were two logical candidates for local metadata standards because the Libraries used DC in all past digital projects. Also, since 2003 most projects have followed the guidelines provided in the CDP DC Metadata Best Practices. ETD-MS, a specific DC application for electronic theses and dissertations, was examined because it provides fuller, customized descriptive metadata for this type of material and was chosen as the standard for CSU’s new ETD initiative. VRA Core and Learning Object Metadata were included to address the emerging local needs of providing access to multi-media digital objects. NISO Z39.87 and PREMIS were targeted for identifying technical metadata requirements for digital still images and other media.

At the time of the investigation, the Libraries had the most experience with DC and the CDP DC Metadata Best Practices. The Task Force believed that a convergence of these standards and schemes would result in the best local metadata scheme to address current and emerging metadata needs. Both DC and the CDP DC Metadata Best Practices facilitated output of simple Dublin Core for OAI Harvesting. Most existing metadata standards applied by the library community build upon Dublin Core, including ETD-MS and VRA Core. The Task Force decided to omit subject-specific metadata fields, because the purpose of this phase was to establish a cross-collection baseline metadata standard that would be expandable according to specific project or disciplinary requirements.

2) Recommend metadata standards and schemes for local best practices, and identify technology developments and changes in use that are necessary to support the best practices.

The Task Force reviewed and compared the definitions and usage of each metadata element in the above-mentioned documents. Based on this research, the Task Force developed a recommendation for a set of core metadata elements for local use (Metadata Best Practices Task Force [MBPTF], 2007b). To reach this recommendation, the Task Force met weekly to discuss findings and progress. In-depth discussions ensued when deciding the use of controlled or uncontrolled vocabularies in the Subject field, Date.Original and Date.Digital elements, refinements of the DC Relation element, and technical metadata elements. The Task Force also designed a unique Identifier element that supported the logical collection feature in DigiTool.

a) Controlled or uncontrolled vocabularies in the Subject field: The Task Force compared *Library of Congress Subject Headings* (LCSH) with specialized controlled vocabularies such as the *Thesaurus of ERIC Descriptors*[8] and the National Agricultural Library *Thesaurus,*[9] and also considered the current trends of user-supplied keywords and full-text searching. The Task Force recognized the value of controlled vocabularies in leading to more accurate subject searches, yet acknowledged the complementary role of user keywords, which provide terms not covered by established vocabularies and which are more familiar to users themselves. Both controlled vocabularies and keywords provide benefits that cannot be replaced by full-text searching, such as aiding in retrieval of more relevant result sets and in user assessment of resources. Thus, in the recommendation, the field is called Subject/Keyword and is mandatory; the content can be either from controlled vocabularies (generic or specialized) or user-supplied keywords.

b) Date.Original and Date.Digital: The qualified DC date element, Date.Created, refers to “date of creation of the resource” (DCMI, 2005). This “resource” is the digital object being described and can be either born-digital or a digitized version of an analog original. For the latter, the information captured in this element is the date an item—for example, a handwritten letter—was digitized, not when the original item was created. The Task Force recognized the importance of having a Date element to capture this information, so, following the CDP DC Metadata Best Practices, recommended a Date.Original element that is not part of qualified DC. While Date.Original was recommended as a “Mandatory if applicable” element, the Date.Digital element, also from the CDP best practices, is mandatory.

c) DC Relation: Dublin Core has thirteen refinements for the Relation element (DCMI, 2005). The Task Force carefully reviewed each refinement's definition and considered its use in the local environment. The final recommendation listed six as "Mandatory if applicable," two as "Recommended," and five as "Optional." See Appendix A for the full listing and the Phase I report for details of the decisions on the use of DC Relation refinements (MBPTF, 2007b).

d) Technical metadata: Reviewing NISO Z39.87 and PREMIS was a fairly long process, because both documents contain numerous data elements and their specifications. However, reviewing these documents increased the Task Force's understanding of current technical and preservation metadata requirements for digital objects. Fortunately, DigiTool, the Libraries' digital repository software, can extract technical metadata which conforms to both NISO and PREMIS. After the Task Force compared DigiTool's technical metadata output against the two standards, it decided to rely on the system's capability and thus did not specify all mandatory technical metadata elements. Previous digital projects had proved that manually gathering this information is too time consuming to sustain.

e) Identifier: All metadata schemas consulted recommend using a unique identifier for each digital object. Both DC and the CDP DC Metadata Best Practices recommend using a string or number conforming to a formal identification system. To avoid potential duplication, and to support future computer manipulation of the Identifier, the CSU Core requires the Identifier to be constructed in a specific manner, consisting of four elements: a four-character collection code, a four-character sub-collection code, a six-digit accession number (which may or may not have meaning), and an optional four-character sub-sub-collection code. This also supports logical collections, a DigiTool feature that allows the Digital Repositories managers to define a pre-coordinated search that targets the 'non-natural' language of the Identifier. The digital content librarian maintains the master list of codes.

3) Seek input from various library groups involved in digital projects and user delivery via presentations and reports available on the staff wiki.

It took roughly three weeks for the Task Force to complete its research and draft its Phase I report. To supplement the report, the Task Force also prepared three scenarios applying core elements to different types of digital objects: a PDF electronic thesis, a PDF faculty paper from the CSU Atmospheric Science Paper series, and a digitized version of a glass plate negative from the University Historic Photographs Collection (MBPTF, 2007b, appendix). On July 12, 2007, the Task Force held two open forums, inviting members from the Digital Repositories Matrix Team, Archives and Special Collections, the CONTENTdm Group, College Liaisons, the Digital Projects Group, and the Copy Cataloging Team. Each forum consisted of a review of the proposed core elements and a discussion of the three scenarios of metadata application. During the discussions, no major issues arose, but minor improvements suggested by the attendees were noted for the Task Force's final recommendation. The Task Force also published all references, working documents, and the final report on the Libraries' internal staff wiki.

4) Present final report to the library management groups for adoption.

The following week, one Task Force member presented the final recommendations for consideration and discussion at a meeting of the Libraries Planning Group (an internal decision-making body). The presentation resulted in approval of the final report, and the core elements have since been officially implemented.

5) Publish final report in the CSU Digital Repository.

The final Phase I report was published in the CSU Digital Repository in July 2008, making it accessible to all CSU faculty and staff, as well as others interested in the topic (MBPTF, 2007b).

Phase II

The work of Phase II built upon Phase I with an objective to develop a data dictionary defining in detail how to implement the core elements locally. To accomplish this task, the Task Force first researched and reviewed existing best practices and data dictionaries available at other institutions and organizations, including:

- *CDL Guidelines for Digital Objects* (California Digital Library, 2007)
- *NC ECHO Dublin Core Implementation Guidelines* (North Carolina ECHO, Exploring Cultural Heritage Online, 2004)
- *Best Practice Guidelines for Digital Collections at University of Maryland Libraries* (Carignan et al., 2007)
- *University of Utah Institutional Repository Metadata Best Practices*, draft version 0.6 (University of Utah Institutional Repository Metadata Subgroup, 2007)
- *Dublin Core Metadata Style Guide* (University of South Carolina Libraries, 2005)
- *A Metadata Framework to Support the Digital Resource Management: User Guidelines for UNTL Metadata Creation*, version 1.2 (University of North Texas, 2004)
- *Metadata Guidelines for Collections using CONTENTdm* (University of Washington Libraries, Metadata Implementation Group, 2004)

These examples gave the Task Force a starting point on which to base the structure and content of its data dictionary. The Task Force decided that for each metadata element, the data dictionary would provide element name, standard referenced, obligation, whether repeatable, definition, comment, refinements, schema, audience, simple DC mapping, input guidelines, and examples (see Appendix B). The Task Force members divided the core elements to work on and later reviewed and combined each individual's work. For the Identifier element, the Task Force also sought input from the members of the Research and Development Services department because of this element's technical applicability in DigiTool. Metadata examples provided in the data dictionary were drawn from past and current local digital projects.

It took roughly two months for the Task Force to complete a first draft of the data dictionary (version 1.0). In September 2007, the Task Force held open forums with relevant library groups to review the recommendation, and, as before, attendee comments and suggestions were noted for the final draft (MBPTF, 2007a).[10] This draft was approved in October 2007 by the Libraries Planning Group and has since been officially implemented.

The Task Force intended the data dictionary to be a guideline document: specific projects may exceed what is outlined or upwardly adjust the obligation of certain metadata elements to fulfill individual project requirements. In these cases, the project manager should create a project-specific data dictionary and note it in the Metadata Schema element in the project metadata. The Task Force also recognized that version 1.0 of the data dictionary may be updated, expanded, or modified as metadata standards/schemes evolve and technologies/systems change.

From June to October 2007, the Task Force successfully completed its first two phases. The results were two guideline documents, the "CSU Core Metadata Elements" and the "CSU Core Data Dictionary" (version 1.0), adopted by the CSU Libraries for metadata creation across all local digital projects. During the research and development process, the Task Force members worked closely and intensively when reviewing standards and composing recommendations; employed the Libraries' internal staff wiki to document its work and progress; actively communicated its work with relevant groups within the Libraries; and sought and incorporated constructive comments into its final recommendations. The final products of the Task Force were closely tied with existing metadata standards/best practices and the Libraries' local digital environment, including collections and technologies. The process not only provided the Task Force members with a wonderful learning experience, but also significantly increased the awareness and understanding of metadata practices and its value among the library faculty and staff, including non-librarian staff working in the technology area.

Next Steps

The Task Force has concluded the Phase IV goal (to investigate and provide a report detailing local human technology resources necessary to comply with recommended best practices) and the Phase III goal (to conduct an inventory of local digital resources and identify how they are currently managed and accessed). This goal will inform the migration of collections currently in CONTENTdm to DigiTool. With this having been accomplished, the initial charge of the Task Force has been completed; however, during the course of this work, the Task Force identified additional remaining tasks:

- Determine the role and suitability of using controlled vocabularies for subject terms, proper names, geographic areas, buildings, etc., including the processes necessary to maintain the vocabularies;
- Develop usability studies to determine if decisions made for the CSU Core Data Dictionary fit the needs of end users;
- Recommend a structure to review metadata plans at the project level, including repurposing MARC records when they exist and migrating existing metadata from other systems; and
- Recommend a process to review, maintain, and update the CSU Core Data Dictionary.

Controlled vocabularies

Controlled vocabularies currently used in the CSU Digital Repository include Library of Congress Subject Headings and personal names, corporate body names, and geographic headings

as established in the Library of Congress Authority File and as used in the CSU Libraries online catalog. Additionally, Inspec thesaurus terms have been added for electrical and computer engineering faculty publications as keywords, which are mapped to the subject search in the repository. The Task Force believes that controlled vocabularies, consistently applied, aid in resource discovery, but also recognizes that the current repository software does not offer an automated way to maintain them. Because manually changing headings will not be sustainable into the future, the Task Force needs to develop strategies for maintaining controlled vocabularies in an automated way. Additionally, the Task Force will need to explore options for incorporating multiple controlled terms in a single search field and the implications for resource discovery.

Usability studies

The Task Force plans to work with the Digital Repositories unit to incorporate testing metadata decisions made for the CSU Core Data Dictionary when it conducts usability testing of the CSU Digital Repository interface. The Task Force needs to determine the type of testing that would evaluate the metadata in terms of both resource discovery and evaluation of a resource according to a user's needs.

Metadata review

The Task Force will recommend a structure for review of metadata planning at the project level. The metadata will conform to the CSU Core Data Dictionary, and the project leader(s) and metadata consultant(s) will determine if additional metadata is required to meet either end-user or system administration requirements. The Inventory Report for pre-existing digital collections held outside the CSU Digital Repository will be examined to determine what changes need to be made to legacy metadata to support repository functionality only, not for exact conformance with the CSU Core.

Maintenance of CSU Core

The Task Force will develop a recommendation for review of the CSU Core. At the minimum, the CSU Core will need to be reviewed following the usability studies mentioned above. Additionally, the Task Force would like to develop a review mechanism that would allow the CSU Core to grow to accommodate not-yet-incorporated metadata standards and elements (e.g., metadata for Geographic Information Systems, earth sciences, and geographic coordinate referencing). Potentially, the Digital Repository will contain data sets, computer software, and other types of digital assets with metadata needs differing from those of text and images, and the CSU Core will need to address those needs.

Conclusion

The process of developing a core data dictionary for the Colorado State University Libraries has successfully resulted in both anticipated and unanticipated outcomes.

When the work began, the Task Force expected to make the planning for digital projects more streamlined. The CSU Core has accomplished this goal by articulating what is mandatory and what is recommended, what decisions should be made at the outset of the project, who should make those decisions, and what potential additional metadata might be applicable to the project. Metadata for projects initiated after the development of the CSU Core has been developed much faster, and more consistently, than was the case for previous projects. Some projects have just used the CSU Core, making decisions only related to identifiers, file naming, structure of compound objects, and the vehicle (MARCXML, DCXML, Excel spreadsheets, etc.) for creating the metadata prior to ingesting into the CSU Digital Repository. Not only were these decisions made more quickly, but the metadata developers and project managers are more confident that the metadata is consistent with existing metadata in the repository and is in compliance with local best practices.

The CSU Core has also provided a framework for training new metadata librarians and staff. Reviewing the data dictionary and understanding the background of how decisions about each element were made now form the core of metadata training. One librarian has completed this review and will soon design his first project-specific data dictionary with minimal guidance.

This streamlined process has contributed to building the repository and creating metadata in a more efficient manner. By having consistency across digital projects, with decisions made at the outset, more employees are involved in metadata creation. As a result, multiple projects are ongoing, and the time and energy of the metadata librarians are spent in managing the production of metadata, not in the actual creation of it. With minimal training, employees having no previous digital project experience are creating original metadata for massive projects, such as the University Historical Photograph Collection, which consists of thousands of CSU images from the late 1800s to the present. They are also extracting MARC records from the Libraries' catalog and manipulating them to create MARCXML for faculty papers and other university publications, as well as using XML editing software to create DCXML on an item-by-item basis for individual resources. These metadata records are batch loaded with the digital objects into the CSU Digital Repository, resulting in large collections of previously unavailable materials readily accessible to the CSU community and the world at large.

The process used to inform employees in various library departments helped communicate the need for quality, consistent metadata, as well as the need to provide this metadata in a cost-efficient manner. By listening to and incorporating comments and suggestions, the Task Force has involved others in the larger picture of building the Digital Repository. This was not entirely unintended, but the level of participation and the resulting buy-in was greater than it had ever been for any previous digital project.

The Task Force anticipates that the effort to build the CSU Core will facilitate discussion of metadata outside the Libraries with other units at CSU. The Digital Repository will be most successful with the support of faculty, students, staff, and administrators of CSU and contributions of their work. The core set of metadata demonstrates the Libraries' commitment to provide a structure and process to enable the discovery, management, display, and archiving of the university's intellectual assets.

Based on a positive experience, the Metadata Best Practices Task Force at CSU would encourage other libraries to utilize a similar approach to develop local best practices and core metadata to meet institutional needs. At times, the work seemed daunting and the scheduled timelines unmanageable. However, the successes include time savings when designing metadata, as well as more efficient involvement of people throughout the library in building the CSU Digital Repository. The process has also encouraged open communication channels and intralibrary cooperation, which provide an excellent foundation for furthering a coordinated metadata management approach.

References

Bastian, D. E. (2005). The Germans from Russia: Documenting the immigrant experience in northern Colorado. *Journal of Archival Organization*, *3*(4), 105-119.

Bishoff, L., & Meagher, E. S. (2004). Building Heritage Colorado: The Colorado digitization experience. In D. I. Hillmann & E. L. Westbrooks (Eds.), *Metadata in practice* (pp. 17-36). Chicago: American Library Association.

Breitbach, B. M., Tracey, R., & Neely, T. Y. (2002). Managing a digital collection: The Garst photographic collection. *Reference Services Review*, *30*(2), 124-142.

California Digital Library. (2007). CDL guidelines for digital objects. Retrieved July 19, 2008, from http://www.cdlib.org/inside/diglib/guidelines/

Caplan, P., & Guenther, R. (2005). Practical preservation: The PREMIS experience. *Library Trends*, *54*(1), 111-124.

Carignan, Y., Evander, J., Gueguen, G., Hanlon, A., Murray, K., Roper, J., et al. (2007). Best practice guidelines for digital collections at University of Maryland Libraries. (2nd ed. draft.) (Susan Schreibman, Ed.) Retrieved July 19, 2008, from http://tomos.umd.edu/dcr/publications/best_practice.pdf

Cherry, T. K. B. (2004). NC ECHO: More than a web portal: A comprehensive access to special collections and digitization program. *North Carolina Libraries (Online)*, *62*(4), 225-234. Retrieved May 6, 2008, from http://www.nclaonline.org/NCL/ncl/NCL_62_4_Winter2004.pdf

Collaborative Digitization Program. (2006). Dublin Core metadata best practices, version 2.1.1. Retrieved July 19, 2008, from http://www.bcr.org/cdp/best/dublin-core-bp.pdf

Dalton, J. (2004/2005). Scholarly articles featuring projects using Dublin Core: A content analysis. *Journal of Internet Cataloging*, *7*(3/4), 59-84.

Dublin Core Metadata Initiative. (2005). Using Dublin Core – Dublin Core qualifiers. Retrieved July 19, 2008, from http://dublincore.org/documents/usageguide/qualifiers.shtml

Dublin Core Metadata Initiative. (2008). Dublin Core metadata element set, version 1.1. Retrieved July 19, 2008, from http://dublincore.org/documents/dces/

Elings, M. W., & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across cultural heritage communities. *Visual Resources Association Bulletin*, *34*(1), 7-14.

Foulonneau, M. (2007). Information redundancy across metadata collections. *Information Processing & Management*, *43*(3), 740-751.

Goldsmith, B., & Knudson, F. (2006). Repository librarian and the next crusade: The search for a common standard for digital repository metadata. *D-Lib Magazine*, *12*(9). doi:10.1045/september2006-goldsmith

Guerard, G., & Chandler, R. L. (2006). California Cultures: Implementing a model for virtual collections. *Journal of Archival Organization*, *4*(1/2), 45-67.

Hicks, E. A., Perkins, J., & Maurer, M. B. (2007). Application profile development for consortial digital libraries. *Library Resources & Technical Services*, *51*(2), 116-131.

Institute of Electrical and Electronics Engineers, Inc. (2002). Draft standard for learning object metadata. Retrieved July 19, 2008, from http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

Lange, H. R., Gravdahl, J., & Leech, H. (1994). The physical characteristics of poster art: Seeking an appropriate list of descriptors. *Visual Resources*, *X*, 235-250.

LeBlanc, J., & Kurth, M. (2008). An operational model for library metadata maintenance. *Library Resources & Technical Services*, *52*(1), 54-59.

Liu, S., & Meyer, L. M. (2008). Carnations and the floriculture industry: Documenting the cultivation and marketing of flowers in Colorado. *Journal of Archival Organization, 6*(1-2), 6-23.

Lopatin, L. (2006). Library digitization projects, issues and guidelines. *Library Hi Tech*, *24*(2), 273-289.

Lourdi, I., Papatheodorou, C., & Nikolaidou, M. (2007). A multi-layer metadata schema for digital folklore collections. *Journal of Information Science*, *33*(2), 197-213.

Ma, J. (2006). Managing metadata for digital projects. *Library Collections, Acquisitions, & Technical Services*, *30*(1), 3-17.

Medeiros, N. (2006). On the Dublin Core front: Metadata in a global world. *OCLC Systems & Services*, *22*(2), 89-91.

Metadata Best Practices Task Force. (2007a). CSU core data dictionary, version 1.0. Fort Collins: Colorado State University Libraries. Retrieved July 19, 2008, from http://hdl.handle.net/10217/3150

Metadata Best Practices Task Force. (2007b). Phase One report: Core metadata elements for CSU electronic theses and dissertations, faculty papers, and University Historic Photographs Collection (glass plate negatives). Fort Collins: Colorado State University Libraries. Retrieved July 19, 2008, from http://hdl.handle.net/10217/3146

National Information Standards Organization. (2006). Data dictionary – technical metadata for digital still images. Retrieved July 19, 2008, from http://www.niso.org/kst/reports/standards/

Nelson, C., & Rutstein, J. S. (1995). Posters as a library resource: The International Poster Collection at Colorado State University. *Collection Management*, *20*(1-2), 115-138.

Networked Digital Library of Theses and Dissertations. (2001). ETD-MS: An interoperability metadata standard for electronic theses and dissertations, version 1.00, revision 2. Retrieved July 19, 2008, from http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html

North Carolina ECHO, Exploring Cultural Heritage Online. (2004). NC ECHO Dublin Core implementation guidelines. Retrieved July 19, 2008, from http://www.ncecho.org/ncdc/ncdublincore.htm

PREMIS Working Group. (2005). Data dictionary for preservation metadata, version 1.0. Retrieved July 19, 2008, from http://www.oclc.org/research/projects/pmwg/premis-final.pdf

Rettig, P. J. (2001). A web of posters: The International Poster Collection at Colorado State University Libraries. *Colorado Libraries*, *27*(3), 18-21.

Rettig, P. J. (2002). Administrative metadata for digital images: A real world application of the NISO draft standard. *Library Collections, Acquisitions, & Technical Services*, *26*(2), 173-179.

Shreeves, S. L., Riley, J., & Milewicz, L. (2006). Moving towards shareable metadata. *First Monday*, *11*(8). Retrieved July 19, 2008, from http://www.firstmonday.org/issues/issue11_8/shreeves/index.html

Simeoni, F., Yakici, M., Neely, S., & Crestani, F. (2008). Metadata harvesting for content-based distributed information retrieval. *Journal of the American Society for Information Science & Technology*, *59*(1), 12-24.

Stvilia, B., & Gasser, L. (2008). Value-based metadata quality assessment. *Library & Information Science Research*, *30*(1), 67-74.

University of North Texas. (2004). A metadata framework to support the digital resource management: user guidelines for UNTL metadata creation, version 1.2. Retrieved July 19, 2008, from http://texashistory.unt.edu/guides/help/UNTL-Metadata-Guide.htm

University of South Carolina Libraries. (2005). Dublin Core metadata style guide. Retrieved July 19, 2008, from http://www.sc.edu/library/digital/dacmetadata.html

University of Utah Institutional Repository Metadata Subgroup. (2007). University of Utah institutional repository metadata best practices, draft version 0.6. Retrieved July 19, 2008, from http://uspace.utah.edu/materials/IRmetadataBPv0p6.doc

University of Washington Libraries, Metadata Implementation Group. (2004). Metadata guidelines for collections using CONTENTdm. Retrieved July 19, 2008, from http://www.lib.washington.edu/msd/mig/advice/default.html

Visual Resources Association. (2007). VRA Core 4.0. Retrieved July 19, 2008, from http://www.vraweb.org/projects/vracore4/

Notes

1. See the CSU Libraries' digital collections at http://digital.library.colostate.edu/.
2. For more background on the International Poster Collection and its metadata, including use of the NISO draft standard, see Lange, Gravdahl, and Leech (1994), Nelson and Rutstein (1995), Rettig (2001), and Rettig (2002). The International Poster Collection website is at http://lib.colostate.edu/posters/.
3. For additional background on the Garst Photographic Collection, see Breitbach, Tracey, and Neely (2002). The Garst Photographic Collection website is at http://lib.colostate.edu/wildlife/.
4. For more on CONTENTdm, see http://contentdm.com/.
5. For additional information on the Germans from Russia Collection, see Bastian (2005). The website is at http://lib.colostate.edu/gfr/. See Liu and Meyer (2008) for more information about the AgNIC (Agriculture Network Information Center) Carnations and the Floriculture Industry Collection. The website is at http://lib.colostate.edu/archives/agriculture/carnations/.
6. For more on DigiTool, see http://www.exlibrisgroup.com/category/DigiToolOverview.
7. "Metadata Best Practices Task Force," by Carmel Bush, sent to the Task Force via email June 19, 2007.
8. ERIC is the Education Resources Information Center. See http://www.eric.ed.gov/ERICWebPortal/resources/html/help/help_popup_thesaurus.html.
9. See http://agclass.nal.usda.gov/agt.shtml.
10. See CSU Data Dictionary (version 1.0) at http://hdl.handle.net/10217/3150. At the time of writing this article, the CSU Core Data Dictionary has been updated to version 1.1. See http://hdl.handle.net/10217/3147.

Appendix A: Summary Table of Core Elements

Element Name	Repeatable	Obligation
Identifier	Yes	Mandatory
Title	No	Mandatory
Date.Digital	No	Mandatory
Publisher	Yes	Mandatory
Rights	Yes	Mandatory
Subject/Keyword	Yes	Mandatory
Type	Yes	Mandatory (extracted in DigiTool)
Format	No	Mandatory (extracted in DigiTool)
Format.Extent	Yes	Mandatory (extracted in DigiTool)
Additional Technical Metadata	Yes	Mandatory (extracted in DigiTool)
MetadataSchema	Yes	Mandatory
Date.Original	No	Mandatory if Applicable
Creator	Yes	Mandatory if Applicable
Title.Alternative	Yes	Mandatory if Applicable
Language	Yes	Mandatory if Applicable
Thesis.Degree.Name	No	Mandatory if Applicable
Thesis.Degree.Level	No	Mandatory if Applicable
Thesis.Degree.Discipline	No	Mandatory if Applicable
Thesis.Degree.Grantor	No	Mandatory if Applicable
Description.Abstract	Yes	Mandatory if Applicable
Source	Yes	Mandatory if Applicable
Relation.IsFormatOf	Yes	Mandatory if Applicable
Relation.IsPartOf	Yes	Mandatory if Applicable
Relation.HasPart	Yes	Mandatory if Applicable
Relation.IsVersionOf	Yes	Mandatory if Applicable
Relation.IsReplacedBy	No	Mandatory if Applicable
Relation.Replaces	No	Mandatory if Applicable
Relation.IsRequiredBy	Yes	Recommended
Relation.Requires	Yes	Recommended
Coverage.Spatial	Yes	Recommended
Coverage.Temporal	Yes	Recommended
Description	Yes	Recommended
Contributor	Yes	Recommended
Contributor.Role	Yes	Optional
Relation.ConformsTo	Yes	Optional
Relation.HasVersion	Yes	Optional
Relation.References	Yes	Optional
Relation.IsReferencedBy	Yes	Optional
Audience	Yes	Optional
Description.TableOfContents	No	Optional
Relation.HasFormat	Yes	Optional

Appendix B: Data Dictionary Structure

The CSU Core Data Dictionary provides the following attributes for the metadata elements

Element Attribute	Description
Element Name	The unique name that identifies the element.
Standard Referenced	The metadata standard(s) consulted that served as a model for the element. One or more of the following: o *Dublin Core Metadata Initiative* (DC) o *Collaborative Digitization Program Dublin Core Metadata Best Practices* (CDP) o *Networked Library of Digital Theses and Dissertations ETD-MS - an Interoperability Metadata Standard for Electronic Theses and Dissertations* (ETD-MS) o *Visual Resources Association VRA Core* (VRA Core) o *Institute of Electrical and Electronics Engineers Learning Object Metadata Standard* (LOMS) o *NISO/ANSI Z39.87: Technical Metadata for Digital Still Images* (Z39.87) o *PREMIS Data Dictionary* (PREMIS)
Obligation	States whether the element is: o Mandatory o Mandatory if applicable o Recommended o Optional
Repeatable	States whether the element may be repeated: o Repeatable o Non-repeatable
Definition	A statement that represents the concept and essential nature of the term.
Comment	Additional information about the term or its application as applied in the CSU context.
Refinements	Lists valid qualifiers for the element.
Schema	Lists valid schema to be used in the element.
Audience	Lists the intended audience for the element: o System o Manager (curator, repository manager) o Staff User o End User
Simple DC Mapping	The simple Dublin Core to which this element maps for metadata sharing via OAI harvesting
Input Guidelines	Provides guidance about entering and encoding values for the element and its refinements.
Examples	Instances of how the element is used.