



SPEC Kit 298

Metadata
July 2007

Jin Ma

Catalog/Metadata Librarian

Baruch College, The City University of New York



ASSOCIATION OF RESEARCH LIBRARIES

Series Editor: Lee Anne George

SPEC Kits are published by the

Association of Research Libraries
21 Dupont Circle, NW, Suite 800
Washington, DC 20036-1118
P (202) 296-2296 F (202) 872-0884
[http://www.arl.org/spec/
pubs@arl.org](http://www.arl.org/spec/pubs@arl.org)


ISSN 0160 3582

ISBN 1-59407-791-6

978-1-59407-791-3

Copyright © 2007

This compilation is copyrighted by the Association of Research Libraries. ARL grants blanket permission to reproduce and distribute copies of this work for nonprofit, educational, or library purposes, provided that copies are distributed at or below cost and that ARL, the source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the US Copyright Act.

 The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (R1997) Permanence of Paper for Publications and Documents in Libraries and Archives.

SURVEY RESULTS

Executive Summary	11
Survey Questions and Responses	17
Responding Institutions	61

REPRESENTATIVE DOCUMENTS

Organization Charts

Boston College	
Boston College Libraries Organization Chart	66
University of California, Los Angeles	
UCLA Library Cataloging and Metadata Center	67
Cornell University	
Discovery Systems and Services	68
University of Florida	
Cataloging and Metadata Department	69
Indiana University	
IU Digital Library Program	70
Library of Congress	
Organization Chart	71
Massachusetts Institute of Technology	
Cataloging and Metadata Services.....	72
Ohio State University	
The Ohio State University Libraries Organizational Chart	73
Pennsylvania State University	
PennState University Libraries Organization Chart	74
University of Virginia	
Cataloging Services	75

University of Washington	
Resource Acquisition and Description.....	76
Washington State University	
Washington State University Libraries. All Libraries. Organization Chart.....	78

Position Descriptions

University of Alberta	
Metadata and Cataloguing Librarian.....	80
Boston College	
Digital Resources Cataloger.....	82
University of California, Davis	
Electronic Resources Librarian	84
Case Western Reserve University	
Head of Bibliographic/Metadata Services	86
Duke University	
Technical Services Archivist/Encoding Specialist	89
Duke University	
Metadata Architect/Programmer	92
Indiana University	
Metadata Librarian	94
Johns Hopkins University	
Metadata Librarian	97
Massachusetts Institute of Technology	
Metadata Librarian	98
University of North Carolina at Chapel Hill	
Electronic Resources Cataloger	99
Pennsylvania State University	
Metadata Librarian	101

Mission Statements and Group Charges

Cornell University	
Metadata Services	104
North Carolina State University	
Metadata Unit	106
University of Virginia	
Metadata Steering Group	107
University of Washington	
Metadata Implementation Group.....	109

Policies, Procedures, and Guidelines

Brigham Young University	
Metadata Standards.....	112
BYU Minimal Standard Metadata Crosswalk.....	113
University of California, Irvine	
California Digital Library Guidelines for Digital Objects	114
University of Chicago	
Guidelines for Use of Technical Metadata: Technical Metadata Dictionary	128
University of Florida	
METS Files for Digital Resources in UFDC.....	137
University of North Carolina at Chapel Hill	
Data Dictionary	152
North Carolina State University	
Course Catalogs Project Metadata Creation Guide.....	156
University of Virginia	
GDMS (General Descriptive Modeling Scheme).....	159
Digital Library Implementation–Generic Text TEI Content Model.....	160
University of Washington	
Metadata Guidelines for Collections Using CONTENTdm	163

SELECTED RESOURCES

Books	167
Articles	168
Web Resources	170



SURVEY RESULTS

EXECUTIVE SUMMARY

Introduction

Metadata is often called “data about data.” It has been used by various communities creating geo-spatial data, social and scientific datasets, enterprise applications, data warehouses, educational resources, and bibliographic data. In the traditional library world, catalog records are metadata, as they contain information about the library’s collection of “data,” i.e., the books and journals that make up its collections. Increasingly, libraries have been adopting emerging metadata standards such as Dublin Core, EAD, MODS, and TEI to describe, discover, preserve, manage, and provide access to electronic resources and digital objects. This is accomplished through three types of metadata: descriptive metadata that describes the intellectual content of the object; structural metadata that ties each object to others to make up logical units; and administrative metadata that manages the object or controls access to it.

This SPEC survey investigated how metadata is implemented in ARL member libraries: which staff are creating metadata and for what kinds of digital objects, what schemas and tools they use to create and manage metadata, what skills metadata staff need and how they acquire them, and the organizational changes and challenges that metadata has brought to libraries.

Background

This survey was distributed to the 123 ARL member libraries in February 2007. Sixty-eight libraries (55%) responded to the survey, of which 67 (99%) reported creating metadata for digital objects at their institutions. One respondent started as early as 1989 and five followed in the subsequent five years. The first sharp increase occurred in 1995 and 1996, when 11 additional libraries began metadata activities. This increase coincided with the creation of the Dublin Core metadata standard at a March 1995 invitational workshop held in Dublin, Ohio. Between 1998 and 2001, 30 more libraries began creating metadata. The activity reached a peak at the turn of the millennium, with 10 libraries entering the metadata arena in 2000. Another peak in 2003, with nine start-ups, followed the availability of DSpace and other institutional repository software. The final five start-ups began between 2004 and 2007.

Metadata Projects and Practices

The primary factor driving the creation of metadata is the responding libraries’ involvement in digitization projects (66 of 67 responses or 99%). Metadata also plays an important role in institutional repositories (54%). Other initiatives and projects that have promoted the use of metadata are: Web content management, datasets, subject-based and edu-

cational repositories, metadata registries, digital media labs, EAD-finding aids, and online journal publishing. As one respondent commented, metadata is distributed throughout several parts of the library and is more broadly applied than solely to digitization projects. Consequently, metadata has been created to describe and provide access to a wide variety of digital resources, including images, text, collections, audio, maps, video, datasets, EAD finding aids, theses, and Web pages.

Metadata Standards

The metadata schemas most widely used by survey respondents are MARC (91%), Encoded Archival Description (84%), Dublin Core (78%), and Qualified Dublin Core (67%). Other commonly used schemas include Text Encoding Initiative Header, Metadata Object Description Schema, and Visual Resources Association Core Categories. A few respondents reported using an array of other schemas for geospatial data, learning objects, works of art, MPEG multimedia files, statistics, databases, etc. Some respondents commented that local or “home grown” metadata standards have been developed.

Survey respondents apply a wide range of controlled vocabularies to metadata, including thesauri, indexes, subject headings, authority files, terms, and ontologies. More than half of the responding libraries use LCSH, LC Name Authority File, and Art and Architecture Thesaurus. A significant number use the LC Thesaurus for Graphical Materials I and II, Getty Thesaurus of Geographic Names, and Getty Union List of Artist Names. About a quarter use MeSH and the Geographic Names Information Service. As with schemas, there are a number of other controlled vocabularies in use, including locally created ones.

Metadata Creation and Management

When asked whether metadata is created manually or automatically, all but one respondent reported

that metadata is created manually. Nine of these also create metadata automatically and 16 also create metadata automatically with human intervention. Eighteen of the respondents reported using all three methods.

The majority of respondents has multiple metadata creators, primarily catalogers (87%), archivists (72%), metadata librarians/specialists (59%), and subject librarians/specialists (49%). Support staff (66%) and student workers (57%) are important contributors to metadata creation and 42 institutions (62%) reported that content creators provide metadata. Database librarians, programmers, preservation librarians, special collections librarians, curators, digital initiatives librarians, and digital programs librarians also contribute metadata. Given the collaborative nature of metadata-related initiatives and projects, it is not surprising that 35 institutions (52%) have accepted metadata from project partners outside of the libraries and 20 (29%) have accepted metadata from vendors.

Survey respondents identified over two dozen software products and tools that they have used for metadata generation. The most commonly used include spreadsheet software such as Excel, relational databases such as Access, Oracle, and MySQL, and MARCEdit. Many respondents also use XML editors with support for XML editing and validation, schema and DTD editing and validation, and XSL editing and transformation such as Oxygen, XML Spy, Stylus Studio, and XMetaL. Quite a few respondents also listed locally developed tools. Almost everyone uses a combination of products for creating and editing metadata.

In addition to metadata editors and generators, there are various sophisticated digital repository and content management systems in use that support metadata creation, editing, and delivery. Other than locally developed systems, DSpace and

CONTENTdm are by far the most frequently used software. Other commonly used systems include Fedora, Luna Insight, DLXS, and Greenstone. More than a dozen other systems were also identified.

Interoperability is essential to facilitate the exchange and sharing of metadata and to enable cross-domain searching. The survey responses indicate that various attempts have been made to achieve metadata interoperability. Fifty-three respondents (83%) report that they have adopted the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Forty-seven libraries (73%) use metadata crosswalk. Other advanced methods and standards are being used to promote metadata interoperability and management, including METS (45%), RDF (25%), metadata registries (20%), and application profiles (20%).

Metadata Quality Control

Respondents were asked how they maintain quality control for metadata and to briefly describe their quality control methods. Fifty-six libraries (83%) reported that metadata are manually checked and approved before publishing. Forty-one (73%) indicated that metadata created by users or content creators are checked and approved by library staff. One respondent stated that their library checks 10% of in-house created metadata as well as 10% of vendor created metadata. Other quality control methods include authority control, XML and schema/DTD validation, and compliance with application profiles. Some respondents mentioned that they use locally developed scripts or a variety of open-source and commercial quality control software.

The comments indicate that different quality control measures are used for different projects. Some believe that more and more content creators will create metadata, which will need more efforts on quality control. One respondent mentioned that they are "currently investigating more automated

methods of metadata checking. This is especially important for content creators." Some commented that metadata creation is time-consuming and expensive; another that the challenge is to reconcile metadata quality vs. metadata cost.

Organizational Change

Fifty-five libraries (85%) reported organizational changes in response to the demands of metadata services while ten reported no organizational changes. Existing positions were redefined to incorporate metadata responsibilities at 36 libraries (62%). Twenty-six institutions (45%) created at least one new metadata position; twelve of these positions were given primary responsibility for managing metadata activities. A variety of titles are used, some of which include the term "metadata," for example: "Metadata Librarian," "Metadata Specialist," "Catalog/Metadata Librarian," and "Metadata Architect." Other titles are: "Text Processing and Mark-up Coordinator," "Digital Projects Coordinator," "Digital Collections Librarian," "Digital Content Librarian," "Digital Services Librarian," "Digital Projects Archivist," and "Electronic Resources Librarian."

Seven separate new units for metadata services were created with the names "Metadata Unit," (two responses) "Metadata Services," "Quality Control Unit," "Digital Access," "Digital Resources Metadata Section," and "Cataloging and Metadata Services." Thirteen respondents incorporated metadata services into existing departments and renamed them. For example, "Cataloging Services" became "Cataloging and Metadata Services;" "Special Collection Team" was renamed "Special Collections and Metadata Section;" and "Access, Support, and Accounting" changed to "Scholarly Resources Integration Department." A larger number of respondents (21 or 36%) incorporated metadata services into existing departments without making any name changes.

About half of the respondents reported that metadata activities are distributed across several departments of the library. Several libraries created temporary term positions to provide additional assistance. A few libraries are in the planning stages of reorganizing to accommodate metadata activities.

Metadata Staffing

Nineteen libraries reported that metadata librarians have primary responsibilities for the management and coordination of metadata activities in their organizations. Another 19 answered “Other librarian;” many of them are the heads of units such as Cataloging, Digital Programs, and Library Technology. At ten libraries, a metadata team/committee/working group plays the leadership role. Archivists play a primary role at three libraries.

Survey respondents were asked the number of full-time and part-time positions and total FTEs for ten different categories of staff who contribute to metadata-related services. Forty-five respondents reported they have staff working full-time on metadata activities, most commonly in the positions of metadata librarian, cataloger, and support-staff, followed closely by programmer and archivist. Fifty-two respondents have staff working part-time on metadata activities. The top four part-time positions are cataloger, archivist, student worker, and support staff. There are significantly more individuals involved on a part-time basis (a total of 583 reported) than full-time (349 individuals). These 932 individuals spend the equivalent to 521.24 hours on metadata activities.

Thirty of 61 responding libraries employ between one and eight individuals, both part-time and full-time, for metadata-related activities; their total is 148 individuals at 84.5 FTE. The average is 5 individuals and 2.82 FTE. Thirty others employ between ten and 68 individuals for a total of 641 indi-

viduals at 368 FTE. Their average is 21 individuals and 12.26 FTE. The remaining library employs 143 metadata staff at 69 FTE.

Not surprisingly, since more than half of the libraries reported that metadata activities are distributed across the institution, most respondents rely on a wide variety of staff to cover metadata operations. Some recurring combinations of staff include metadata librarian(s), cataloger(s), programmer(s), and support staff; archivist(s), programmer(s), support staff, and students; cataloger(s) and support staff; and metadata librarian(s), cataloger(s), archivist(s), and support staff. The libraries that are managing digitization projects, digital repositories, data sets, and Web content have the largest number of staff and the widest range of staff categories.

Metadata Staff Training

On-the-job training, library school, and professional association-sponsored workshops are among the top three sources from which metadata staff received their initial metadata training. The majority of the respondents reported using the following opportunities to keep up-to-date on metadata knowledge and skills: electronic discussion lists, professional journals and readings, conferences, discussion with peers, blogs and online readings, on-the-job training, and professional association-sponsored workshops. Twenty-four of the responding libraries hold in-house workshops for initial staff training and 19 of these also hold workshops for keeping staff up-to-date. An additional nine respondents hold workshops to keep staff up-to-date. Software documentation, subscription-based online tutorials, and consulting with other institutions provide other learning opportunities.

Metadata staff members from the responding libraries attend a wide variety of international, national, regional, and local conferences and workshops on metadata-related topics. Conferences

sponsored by professional organizations such as IFLA, ALA (ALCTS, LITA), ARL, and ASIS&T provide ample programs and opportunities for metadata staff. The annual Dublin Core conferences bring together leading metadata researchers and professionals from around the world. Metadata is a frequently discussed topic at digital libraries conferences including the Joint Conference on Digital Libraries, Digital Library Federation Forums, and International Conferences on Open Repositories. OCLC regional networks, regional library associations, and consortia provide educational opportunities for staff at all levels.

Metadata Librarian Qualifications and Responsibilities

Forty-eight of the responding libraries have at least one metadata librarian position; 42 of these require an MLS degree. Knowledge of emerging metadata standards and experience with MARC cataloging are required by all but a few libraries. Soft skills such as communication skills, problem-solving skills, and ability to work cooperatively and independently are also required by over 70% of respondents. About one third of the responding libraries require advanced knowledge of metadata crosswalks, interoperability, and experience with integrated library systems. Experience with institutional repositories and digital content management systems, and knowledge of XML and OAI are listed as desirable qualifications by about half of the respondents.

The survey responses indicate that at most of the responding institutions, the metadata librarian plays a leadership role in metadata activities while performing the following functions: consulting on metadata options in terms of metadata standards; working with systems personnel, subject specialists, project partners, and even end-users on metadata-related issues; documenting metadata policies, procedures, and guidelines; and training staff.

Metadata Challenges

The survey respondents were asked to list the top three metadata challenges facing their libraries. Many respondents face the challenge of implementing organizational changes. Typical comments identified the challenges of “creating the right internal organization for providing metadata services,” “developing/accommodating workflow for metadata creation,” “developing workflow between departments,” and “managing projects that cross so many departments/divisions of the libraries and that involve other units across campus.”

The proliferation of emerging standards poses the challenges of reaching decisions on metadata standards to use for various projects, as well as implementing consistent standards and tools and keeping up-to-date on emerging standards. Consequently, interoperability becomes the most critical issue. Survey respondents commented on the lack of “system-wide infrastructure,” the lack of “system interoperability,” and the difficulty of “interoperating existing online collections with new collections which are supported by other platforms,” and the “cost and difficulty of creating good, consistent metadata across multiple formats and implementation systems.”

Another common challenge is the lack of staff and resources for metadata creation and management. Likewise, some libraries had difficulties “finding competent people to do specific tasks,” “training new personnel,” and “retaining personnel once trained.” Quite a few respondents urged automating metadata creation as much as possible.

Other challenges that survey respondents mentioned include authority control, preservation of metadata along with digital objects, sustainability and scalability of metadata creation, and “meeting increasing demand for metadata skills and support throughout the library.”

Conclusion

The survey responses provide a general overview of the current state of metadata implementation in ARL member libraries. Comments indicate that many libraries are in a period of transition as they attempt to determine the best organization, staffing, and system infrastructure. Metadata involves staff from many different units and many respon-

dents are implementing appropriate workflow, streamlining metadata production, and developing policies and guidelines for best practices. As the survey data reveal, metadata interoperability is among the top challenges that libraries face due to the proliferation of standards, tools, and options available in a rapidly changing environment where infrastructure and rules are not well established.