

DLF-Aquifer Metadata Institutional Survey Report

Aquifer Metadata Working Group
Report on the Institutional Metadata
Survey Results

Version 1.0

Last revised March 16, 2007

Table of Contents

Executive Summary	3
Methodology	5
Online Survey of DLF Institutions.....	5
Analyzing Results.....	6
Results	8
Section 1: Descriptive Metadata Formats.....	8
Section 2: Use of MODS	9
Section 3: Metadata Tools & Processes.....	10
Appendix A: Survey of MODS Use & Metadata Tools.....	11
Introduction.....	11
Section 1 of 5: Descriptive Metadata Formats	11
Section 2 of 5: Use OF MODS	12
Section 3 of 5: Metadata Tools and Processes	13
Section 4 of 5: Open Questions and Comments	13
Section 5 of 5: Institutional Information.....	13
Thanks!.....	14
Appendix B: Survey Cover Letter.....	15
Appendix C: Qualitative Data from Survey.....	17
Section 1 of 5: Descriptive Metadata Formats	17
Section 3 of 5: Metadata Tools & Processes.....	17
Section 4 of 5: Open Questions and Comments	20

Executive Summary

The Digital Library Federation (DLF) Aquifer Metadata Working Group (MWG) was charged with recommending metadata policies and best practices to the DLF Aquifer Implementation Group and designing workflows for metadata harvesting, creation and enhancement. As a step towards meeting these goals, the group has undertaken a survey of DLF institutions in order to discover what metadata formats are currently used among DLF institutions to create descriptive metadata within American cultural heritage collections, to gauge the current use of MODS (as a richer and shareable metadata format), and to assess what tools are used and desired by institutions in their work with metadata. Several key findings emerged from this survey:

- **DLF institutions use a variety of descriptive metadata formats in managing their digital resources locally.** Not surprisingly, format- and system-based factors have a strong impact on local decisions about metadata formats. The focus of DLF Aquifer on improved richness of *shareable* descriptive metadata being made available by DLF institutions for aggregation is an important factor to reinforce.
- **MODS is not widely used as a descriptive metadata format in local content management systems.** Articulating the desired usage of MODS as a richer format than Dublin Core for *sharing* metadata for aggregation, and providing tools that facilitate transformation into MODS of shareable metadata records from native content management system metadata formats will be important for DLF Aquifer moving forward.
- **DLF institutions use a variety of homegrown, open source, and commercial tools to create, manipulate, manage, and preserve descriptive and other metadata in local environments.** Local factors will continue to be a primary contributor to local decisions about how to manage metadata for digital resources through its lifecycle. Transformation for *optimizing shareability* of metadata extracted from locally relevant content management environments should remain a key objective for DLF Aquifer.

The results of this survey point to some issues and considerations for the current and future work of DLF Aquifer, and suggest that there are several areas of need among DLF institutions that the Aquifer project could meet. The following are key issues and considerations that will inform the work of the DLF Aquifer Metadata Working Group:

- **Institutions need tools that facilitate more standardized transformations of descriptive metadata from locally relevant formats into records that can be harvested by data aggregation services for a variety of user-specific needs.** The MWG is currently working with the Library of Congress to create a version of the popular MARC-to-MODS stylesheet that produces a shareable MODS record that adheres to the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records, Version 1.0*. This effort will meet the needs of a significant number of survey respondents who are using MARC as a local metadata format. The survey also indicates that the MWG might profitably begin discussions with communities of TEI, VRA Core, and CDWA Lite users around development of standardized transformation stylesheets oriented towards shareability with these communities..

- **Many respondents indicated local work with EAD to encode finding aids, or guides, to archival collections.** The prevalence of EAD usage indicates that the MWG might profitably work with the archival community to articulate a common understanding of best practices for sharing metadata records for these finding aids as digital objects, and to develop a standardized stylesheet for extracting a shareable metadata record for finding aids that adheres to the *DLF/Aquifer Implementation Guidelines for Shareable MODS Records, Version 1.0*.

Methodology

The Metadata Working Group identified a gap in knowledge about how various DLF and DLF Aquifer institutions were managing metadata, what formats were in production or test, and common and desired tools for managing metadata. It was decided that this information could be quickly and comprehensively gathered using an online survey. During the summer of 2005 the group constructed an online survey which was then distributed among all DLF institutions. In conducting these surveys, the Metadata Working Group hoped to discover current metadata practices and formats among DLF institutions.

Online Survey of DLF Institutions

In an effort to better inform its developing activities, the Metadata Working Group initially sought to discover (1) whether institutions were using the MODS metadata format, and how (production or test), (2) what documentation they may have developed regarding MODS use, and (3) what tools they are currently using to work with metadata. Following targeted email queries of metadata librarians in four DLF institutions, the group decided that, given the brevity of the survey and the number of DLF institutions, it would be more feasible to distribute an online survey. This online survey would allow for quick and comprehensive querying of all DLF institutions while simultaneously providing participants with an opportunity to describe metadata management at their institution. A copy of this online survey is provided in Appendix A.

Participants

Our target audience was metadata librarians at DLF institutions. Additionally, we sought out individuals with a history of active participation in DLF functions, who would thus have a sense of the provenance of the Aquifer project and the potential importance of its work for the larger DLF community. Names and contact information were culled from a DLF database of active members, then supplemented by information on digital library initiatives currently being undertaken at DLF institutions and recommendations made by DLF Aquifer working group members.

A copy of the survey URL was included in a cover letter emailed to two individuals at each institution, pre-identified as likely participants for one of the two surveys being distributed by DLF Aquifer. This letter included background information on the DLF Aquifer project, description of the survey, instructions on completing the survey, and a request to pass the survey along to a more appropriate respondent, if deemed necessary. In an effort to control and more easily assess the data, we asked that each institution submit only one institutional response for the survey. A copy of the survey cover letter is included in Appendix B. Note that this cover letter introduced two separate information-gathering surveys that DLF Aquifer conducted. The Services Working Group's (SWG) survey canvassed current and desired use of DLF institutions' digital collections and services, existing and desired support for those services, and usage assessment methods and frequency. A summary of the SWG's survey results are published separately.

Design

The survey was divided into five sections, with the first four sections addressing different aspects of metadata formats and management and the last section requesting personal

and institutional information for use in analyzing the results. The goals of sections were as follows:

- **Section 1: Descriptive Metadata Formats**
Assess metadata formats currently being used to create descriptive metadata within American cultural heritage digital collections.
- **Section 2: Use of MODS**
Assess current use of MODS and documentation developed.
- **Section 3: Metadata Tools & Processes**
Solicit information about tools institutions are currently using with their metadata and for what purposes.
- **Section 4: Open Questions & Comments**
Provide an opportunity for participating institutions to supply other information, ideas, or questions that may not have been covered or fully explored in the survey, as well as to comment on the survey itself.
- **Section 5: Institutional Information**
Gather information on who participated in this survey, to use in analyzing the results and following up on responses if necessary.

As stated previously, the online survey was intended to canvass a broader audience while still allowing for qualitative descriptions of DLF institutions' use of metadata tools. A copy of the text of the online survey, with screen captures, may be found in Appendix A.

Format

The choice of an online survey format offered (1) the possibility of a faster turnaround time than we might reasonably expect with mailed paper questionnaires and (2) already digitized data, which could also decrease our time spent in analysis of the results.

Timeline

On August 3, 2005, URLs to the online survey were emailed to 34 DLF institutions. (A copy of the survey cover letter is available in Appendix B.) Participants were initially given two weeks to complete the online survey. A decision was made to extend the closing date an extra week when it became clear that many institutions could improve their responses by having more time to gather people and information necessary to complete the survey. The survey closed on August 26, 2005.

Analyzing Results

At the end of the survey period, the Metadata Working Group had collected 27 responses to the online survey. Initial passes through the data revealed a few blank surveys. Removing blank responses resulted in 25 total responses to the survey, with an average of 22 responses to each question (since respondents did not always answer every question). Because of the small sample size of this survey, it was decided that, rather than omitting incomplete responses, the number of respondents for each question would be indicated in the analysis.

Cleaning Data

In the interest of sharing survey findings with others without compromising the anonymity of respondents, qualitative data that could be used to identify an individual or institution was either quantified or was amended to remove any identifying information. The following generic descriptions were used in the data to substitute for institution- or individual-identifying information:

GENERIC DESCRIPTION	DATA REPLACED
[institution]	Name of an institution
[project]	Name of a specific grant project or local initiative

In addition, because most of the responding DLF institutions use American English spelling conventions, some words were re-spelled to fit this standard so that non-American institutions could not be easily identified. The anonymized qualitative survey responses are provided in Appendix C of this report. As mentioned previously, responses to questions specifically soliciting personal or institutional information are not provided or analyzed in this report. Additionally, detailed qualitative responses describing tools an institution uses with metadata were also omitted. Because question 5 – “What tools do you use to manage and manipulate metadata (in any format?)” – produced highly detailed accounts of institutions’ metadata technical infrastructure, and because omitting such specifics from the qualitative response would essentially render the information useless, this data is not provided in the Appendix.

Qualitative Data Analysis

The Aquifer MWG survey contained four questions soliciting qualitative responses: either questions with pre-defined response categories with an open-ended "Other, please specify" option or simply open-ended questions. In addition to listing narrative responses to three of these questions in the survey report (see Appendix C for these anonymized qualitative responses), the decision was made to quantify answers to question 5 according to the specific tools used, the category of tools used, and the number of institutions employing them. In the interest of further protecting the anonymity of respondents, tools used by three institutions or less were omitted from the quantitative data.

Results

The survey was divided into five sections, each addressing different aspects of metadata use and management and/or gathering institutional or individual information for administrative use. This section of the report summarizes qualitative and quantitative responses to survey questions. Responses to questions in sections specifically requesting institution- or individual-identifying information are excluded from this report.

Section 1: Descriptive Metadata Formats

The first section of the survey sought information on what formats DLF institutions were currently using to provide descriptive metadata within American digital cultural heritage collections (i.e., those digital collections whose content might fall under the broad topic of "American culture and life"). Among respondents, EAD (Encoded Archival Description) and MARC are the most prevalent metadata formats, closely followed by DC (Dublin Core, either simple or qualified). Figure 1 below indicates the total number of institutions who reported using each metadata format.

**Which of the following do you use as descriptive metadata for digital resources in American cultural heritage collections?
(25 total respondents)**

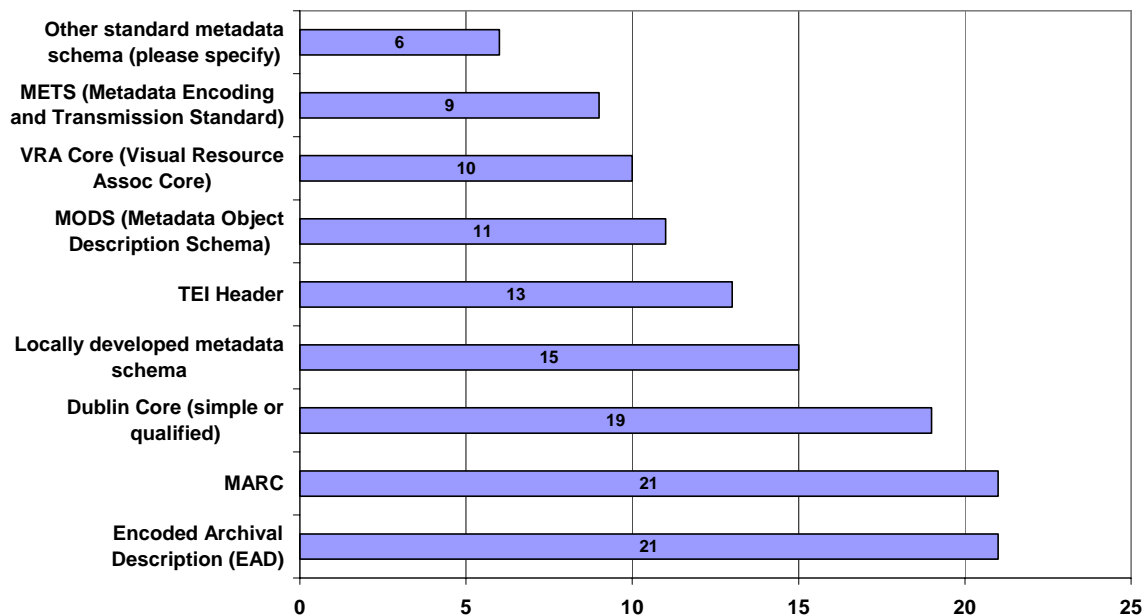


Figure 1: Descriptive Metadata Formats Used

Responses to this question also indicated that DLF institutions employ a variety of formats with their collections: on average, respondents employ five different metadata formats; one-fifth of the responding institutions (5 of 25) use seven or eight different metadata formats with their digital collections. Six respondents provided used the "Other standard

metadata schema (please specify)" portion of this question to provide more specific information that is included in Appendix C. The fact that some of the schemas mentioned are modifications of existing schemas or unique to particular subject areas or disciplines further emphasizes the diverse range of metadata formats currently in use.

Section 2: Use of MODS

As part of its charge for the DLF Aquifer project, the Metadata Working Group recommends metadata policies and best practices. The group identified MODS early on as a format that could support richer descriptions of digital objects and resources, and used this survey to gauge current MODS activity among DLF institutions.

Are you currently using MODS in either a production or test environment? (25 total respondents)

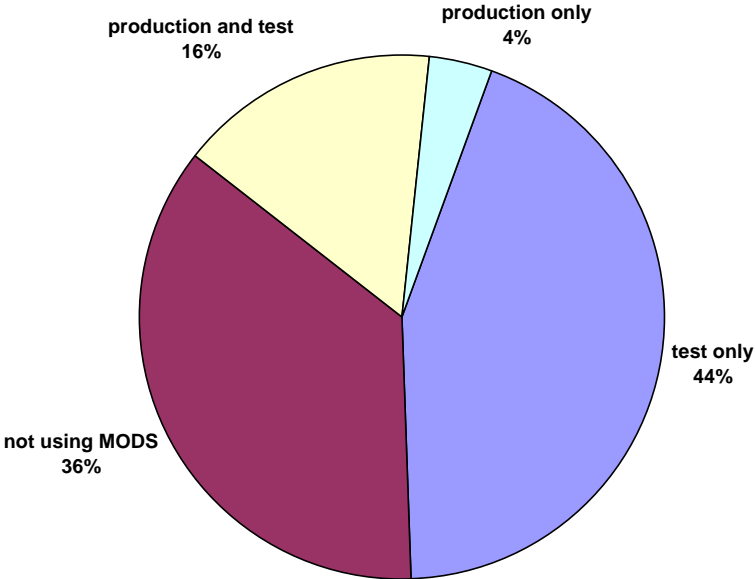


Figure 2: MODS Use Among DLF Institutions

Though MODS is still much less common than other metadata schema (as evidenced in Figure 2), the majority of DLF institutions who responded to this question indicated some use of MODS, whether through testing, production, or both. An important peripheral point to highlight here is the very close relationship between MODS and MARC. The survey, not surprisingly, indicates that a large percentage of respondents (21 of 25) are currently using MARC as one of their descriptive metadata formats for cultural heritage materials. While use of MODS among DLF institutions is still primarily in the testing phase, and few institutions (6 of 25 respondents) have developed any documentation for its use, it is not a stretch to assume that the prevalence of MARC usage indicates some promise for using MODS as a metadata resource to support richer shareable metadata. As indicated by the survey results, a key challenge for the DLF Aquifer Metadata Working Group moving forward will be to engage communities using metadata formats other than MARC in their local content management systems around the transformation of those local records into shareable MODS records as defined in the Aquifer guidelines.

Section 3: Metadata Tools & Processes

Just as there are a range of metadata formats in use to describe digital cultural heritage collections, there are innumerable tools, systems, and scripts currently being used by institutions to create, manipulate, enhance, and transform their metadata. The following table represents a sampling of the 95 specific tools mentioned by 25 respondents.

General Types of Tools Mentioned	Specific Tools Mentions (% of respondents who indicated use)
Integrated Library System	Aleph (8%), SIRSI (12%), Endeavor (4%)
Content Management System	DLXS (12%), Documentum (4%), Vital (4%)
Database	DB2 (4%), EndNote (4%), FileMaker (8%), FoxPro (4%), MS Access (4%), Various flavors of SQL (16%), Oracle (12%), XML databases (12%),
Digital Asset Management System	CONTENTdm (12%), DigiTool (4%), Hyperion (4%), Insight (4%), Verde (4%)
Editor	XMLSpy (24%), BBEdit (4%), DC-dot (4%), Emacs (8%), Inscribe (4%), MARCEdit (16%), NoteTab (16%), OpenOffice (4%), Oxygen (24%), Topologi (8%), vi (4%), XMetaL (16%)
Processor/Extractor	MARC-to-TEI tool (4%), perl scripts (40%), SAXON (16%), Xalan (40%), PHP scripts (12%), JHOVE (8%)

Respondents were forthright and highly descriptive in identifying needed or desired tools for metadata management.

Samples of user responses:

"Software that could transform diacritics and symbols to UTF8 regardless of their form."

"A robust editing environment that would read and interpret XML schemas, and that would allow data entry verification and output that would conform to the specified schema."

"Tools that can allow simple mappings among schema so as to create metadata in a variety of syntax or convert metadata in a variety of schema."

"For projects that include OCR'd text, it would be nice if some of the metadata could be automatically generated."

"In other words, we want an easy way to grab an XML document and edit it in a forms-based environment that requires minimal staff training but produces standards-compliant metadata."

What seems clear from these responses is that institutions will continue to use a variety of tools for creating, managing, and preserving metadata (descriptive and otherwise) in their local content or digital asset management systems. This suggests that continued work by the MWG with a focus on *optimizing shareability* of metadata records extracted from their local contexts and made available for external aggregation is realistic and, in the long run, beneficial to continued exploration of future directions for digital library efforts in academic and research settings.

Appendix A: Survey of MODS Use & Metadata Tools

INTRODUCTION

Background: The DLF Aquifer project is an initiative of the Digital Library Federation to create scalable solutions that enable teaching, learning, and scholarship. Beginning with a collection of digital content in the area of American culture and life, DLF Aquifer will create a test-bed of tools for selecting, collecting and providing access to quality digital content. The Metadata Working Group of the DLF Aquifer project is charged with recommending metadata policies and best practices to the DLF Aquifer Implementation Group. This group will design workflows for metadata harvesting, creation, and enhancement. More information on the DLF Aquifer project is available through the Digital Library Federation's website: <http://www.diglib.org/Aquifer/>.

Purpose of survey: The following brief survey is an attempt to assess what metadata formats DLF institutions are currently using and what tools are in use to manipulate and manage metadata for relevant cultural heritage and humanities oriented digital resources. The goal is to identify those resources and practices which could benefit other participating DLF institutions and the Aquifer project and to have a better understanding of the current metadata landscape. In addition, having identified MODS as a promising format for providing detailed and uniform shareable metadata, the Metadata Working Group is currently developing a profile and recommendations for its use. Input and insight on MODS is highly valued, particularly from member institutions who are already working with MODS. Questions regarding this survey may be directed to the chair of the Aquifer Metadata Working Group, Sarah Shreeves: sshreeve@uiuc.edu.

Completing the survey: This survey should take 15 minutes to complete. There are 9 questions total, divided into 5 sections. Please feel free to consult with others at your institution who could provide useful feedback. This survey is open *through August 17*.

To start the survey, click "Next" (below). To move forward or backward through the survey, click the arrows at the bottom of each screen. If you'd like to leave the survey at any time, just click "Exit this survey" (upper right-hand corner of the screen). Your answers will be saved. In order to resume the survey where you left off, use the same computer to link to the survey URL. Thank you!

SECTION 1 OF 5: DESCRIPTIVE METADATA FORMATS

Various metadata formats may be used to describe digital resources. The Metadata Working Group is interested in those formats currently being used to create descriptive metadata within American cultural heritage digital collections (i.e., those digital collections whose content might fall under the broad topic of "American culture and life").

**1. Which of the following do you use as descriptive metadata for digital resources in American cultural heritage collections?
(check all formats that apply)**

- Dublin Core (simple or qualified)
- Encoded Archival Description (EAD)
- MARC
- METS (Metadata Encoding and Transmission Standard)
- MODS (Metadata Object Description Schema)
- TEI Header
- VRA Core (Visual Resource Assoc Core)
- Locally developed metadata schema
- Other standard metadata schema (please specify)

SECTION 2 OF 5: USE OF MODS

Background: The Metadata Working Group is currently developing a MODS metadata profile for DLF institutions with cultural heritage and humanities material based on the DLF/NSDL OAI Best Practices for Shareable Metadata work. The group is interested in which institutions are already using MODS (either in production workflows or in testbeds) and what insights or resources they would be willing to share with the group to help inform the use of this format. For more information on the DLF/NSDL OAI Best Practices for Shareable Metadata work, please visit their metawiki: <http://oai-best.comm.nsd.org/cgi-bin/wiki.pl>.

2. Are you currently using MODS in either a production or test environment?

- production and test
- production only
- test only
- not using MODS

3. Have developed any documentation for using MODS?

- Yes
- No

4. If you have developed or are planning to develop documentation for using MODS, would it be possible to share this with the Aquifer Metadata Working Group?

Please select your institution from the pull-down menu across from your answer.

Yes	<input type="text"/>
No	<input type="text"/>
N/A	<input type="text"/>

SECTION 3 OF 5: METADATA TOOLS AND PROCESSES

Background: A number of commercial and free tools are available for creating, manipulating, enhancing, and transforming metadata. Some institutions may have also developed their own local resources for working with metadata. Institutions may also work with metadata by using tools provided with their digital library systems.

The following questions attempt to gauge what tools institutions are currently using with their metadata and for what purposes.

5. What tools do you use to manage and manipulate metadata (in any format)? These could be locally developed tools, tools already included in a content management system, tools for transformation of metadata, and so forth. Please list these tools below, and include any comments about their use.

6. Are there tools that you wish you had, which would help you manage and manipulate metadata? If so, please briefly describe.

SECTION 4 OF 5: OPEN QUESTIONS AND COMMENTS

This survey focuses on the immediate concerns of the Aquifer Metadata Working Group. We recognize that there are other issues regarding metadata for cultural heritage digital resources that this survey does not address. If you have additional questions or comments that you feel are relevant to this project, please bring them to our attention by mentioning them in the box below.

SECTION 5 OF 5: INSTITUTIONAL INFORMATION

The following information will be used to determine how many DLF institutions participated in this survey, and to identify those willing to be available for follow-up questions and/or to contribute further information or resources to the project. We appreciate your willingness to complete this survey, and hope we may rely on you to continue to assist us in this DLF endeavor.

8. Name of your institution:

[institution name selected from pull-down menu of DLF institutions]

9. May we contact you for follow-up questions or further information? If so, please enter your name and contact information below.

Name:	<input type="text"/>
Position/Title:	<input type="text"/>
Email:	<input type="text"/>
Phone:	<input type="text"/>

THANKS!

Thank you for your time. We appreciate your feedback. You are free to make changes to your survey response *through August 17*. To make changes to your response, link to the survey URL using the same computer.

If you have any questions, please contact the chair for the Aquifer Metadata Working Group, Sarah Shreeves: sshreeve@uiuc.edu. Information and updates on the DLF Aquifer project are available through the DLF website: <http://www.diglib.org/Aquifer>.

Appendix B: Survey Cover Letter

Greetings from the Digital Library Federation!

As an active DLF institution member, you are being asked to complete the following surveys addressing aspects of digital librarianship at your institution or to forward them to the person or group best qualified to respond. We seek one institutional response per survey.

Survey of MODS Use & Metadata Tools -

<http://www.surveymonkey.com/s.asp?u=469551250385>

Survey of User Services for Digital Collections –

<http://www.surveymonkey.com/s.asp?u=514031250426>

Responses to these surveys will inform the work of DLF Aquifer, an initiative of the DLF. We appreciate the time and consideration you give to our queries, and will share our findings once the survey analysis is complete. Please feel free to consult with others at your institution as you complete the survey. This survey is open through August 17. More information on DLF Aquifer and the surveys is included below.

Thank you for your active interest in the DLF. We look forward to your continued participation in its efforts, and in your prompt and considered response to these surveys.

Sincerely,
Katherine Kott

BACKGROUND:

The Digital Library Federation (DLF) is currently engaged in a collaborative initiative to create scalable tools and services that enable teaching, learning, and scholarship. Called DLF Aquifer, this initiative seeks to improve and expand access to digital library content across institutions, particularly in the humanities and social sciences. More information on DLF Aquifer is available through the DLF website: <http://www.diglib.org/aquifer/>

Survey of MODS Use & Metadata Tools –

<http://www.surveymonkey.com/s.asp?u=469551250385>

PURPOSE: Gather information on metadata formats DLF institutions are currently using, tools used to manipulate and manage metadata, and use of MODS.

PARTICIPANTS: Anyone involved in developing or managing metadata for digital content, either in production or test environments. The URL for the survey may be forwarded to others within your institution who could provide useful feedback.

SCOPE: Nine questions are divided among five areas: 1) Descriptive Metadata Formats; 2) Use of MODS; 3) Metadata Tools & Processes; 4) Open Questions & Comments; and 5) Institutional Information.

CONTACT: Questions regarding this survey may be addressed to the Aquifer Metadata Working Group chair, Sarah Shreeves: <<mailto:sshreeve@uiuc.edu>>sshreeve@uiuc.edu.

Survey of User Services for Digital Collections –

<http://www.surveymonkey.com/s.asp?u=514031250426>

PURPOSE: Gather information on how scholars at DLF institutions currently use digital collections, how that use might be improved, and what is needed in order to improve services.

PARTICIPANTS: Anyone involved in assessing user services for digital collections, assisting users with digital collections and services, or supporting/developing digital collections and services. The URL for the survey may be forwarded to others within your institution who could provide useful feedback.

SCOPE: Fourteen questions are divided among five areas: 1) Assessing Use; 2) Actual & Desired Use; 3) Supporting Digital-Resource Use; 4) Open Questions & Comments; and 5) Institutional Information.

CONTACT: Questions regarding this survey may be addressed to the Aquifer Services Working Group chair, Martin Halbert: <<mailto:mhalber@emory.edu>>mhalber@emory.edu.

Appendix C: Qualitative Data from Survey

The following comments and information were offered in response to the open-ended questions in the survey (e.g., "Other" options at the bottom of a list of pre-formulated responses). Specifically, this Appendix includes qualitative responses to questions 1, 6, and 7 of the survey. The total number of respondents and the number who skipped a question are provided for each question.

In order to preserve the anonymity of respondents, all references to specific institutions, individuals, programs, projects, etc. have been removed from the following raw responses; bracketed ellipses, words, or phrases indicate removed or substituted information. In addition, because most of the responding DLF institutions use American English spelling conventions, words were re-spelled to fit this standard so that non-American institutions could not be easily identified. Answers to questions that specifically queried institution- or individual-identifying information are excluded. Because responses to question 5 essentially provided technical fingerprints for institutions, all responses to that question are likewise omitted from this Appendix.

SECTION 1 OF 5: DESCRIPTIVE METADATA FORMATS

2. Which of the following do you use as descriptive metadata for digital resources in American cultural heritage collections? (check all formats that apply) **(6 responses to this option, 0 skipped question)**

- 1) DDI
- 2) Internally created metadata sets. We've used METS in the creation of a prototype record transfer schema.
- 3) CrossRef metadata for digital journals published by the library (other standards are in use for digital projects that fall outside of "American cultural heritage" definition).
- 4) For VRA Core, we are using a system that uses much of the work/image philosophy of the Core, but it not completely compliant in all places.
- 5) NISO 3987
- 6) DC Terms, FGDC, DDI, MathML

SECTION 3 OF 5: METADATA TOOLS & PROCESSES

6. Are there tools that you wish you had, which would help you manage and manipulate metadata? If so, please briefly describe. **(20 responses, 5 skipped question)**

- 1) Schema-based metadata editor for web environment (possibly Altova Authentic - still under investigation). Collaborative tools for metadata creation/workflow.
- 2) Tool to check digital provenance of image files, Artisia's Teams software

- 3) A tool that allows you to put in a URL or cut & paste a text file, and have it test to see if it's unicode compliant -- and have it show me where any non-compliant entities exist on the page. Also, software that could transform diacritics and symbols to UTF8 regardless of their form.
- 4) - to automatically extract existing descriptive and subject metadata from digital objects
- to automatically generate metadata from digital objects (text, image, sound) - to identify and remove duplicates
- 5) a tool that clearly describes the metadata formatting being used, e.g., utf8 vs. isolatin1
- 6) A robust editing environment that would read and interpret XML schemas, and which would allow data entry verification and output that would conform to the specified schema. The editing environment should include authentication of users and version control. In other words, we want an easy way to grab an XML document and edit it in a forms-based environment that requires minimal staff training but produces standards-compliant metadata.
- 7) Tools that can allow simple mappings among schema so as to create metadata in a variety of syntax or convert metadata in a variety of schema. These tools should be able to validate, parse, and format data with ease.
- 8) Our current tools need to be (and are in the process of being) expanded. For example, we are currently enhancing the Web interface to [local database-driven tool] to allow archivists to work from the Collection (EAD) level down through the individual item (METS) level. Once this enhancement is complete, we will expand [local METS extraction tool] to generate EADs as well as METS object from the [local tool] database. There are other (indeed innumerable) input and processing needs that we haven't addressed yet in the tools. In particular our users need better means for rearranging the hierarchical structure of collections and objects in the course of input. Processing archival collections truly requires a great deal of flexibility so that archivists can experiment and "feel out" the most appropriate structure for a collection. We also need better project management/quality control tools--the means for project managers efficiently to review and correct the metadata after the initial input. And we need a better ways for enabling [local tool] users to tap and making use of existing metadata (from the [digital resource catalog]) Our tools have a long history (some of [local tool] originated even before MOA2, let alone METS and MODS), and because of this they are carrying a lot of baggage. As a matter of expedience we find ourselves adapting old structures and technologies to new ends. Since I think directions in digitization are much clearer now, I for one would welcome an opportunity to rethink and reimplement substantial parts of our metadata gathering and manipulation tools with the clearer ends in mind.
- 9) Need tools that will help manage iterations of schemas as they represent categories of metadata (descriptive, technical for different formats, preservation, etc.), but also complete packages of digital object classes (METS profiles for BookSimple, Images, Journals, etc.). Am investigating SchemaLogic and the full tool set that's part of the Altova suite (XMLSpy).
- 10) -- For projects which include OCR'd text, it would be nice if some of the metadata could be automatically generated; --METS creation/export tool; -- MODS creation/export tool; --Open Source digitization workflow management/metadata system; -- Tools to aggregate digital content from multiple sources, both commercial

and open access -- and to layer on services such as annotation, automatic notification, and incorporation into course management systems.

- 11) We are looking for a new central cataloging tool for images to replace IRIS. We are looking for something that is VRA Core compliant (Core 4 in the fall), that has full authority control functionality, allows us to build complex hierarchy and relationships, sophisticated global change, permissions control, and other functionality we are accustomed to in an ILS. Additionally, for all of our processes, we would like more sophisticated (and targeted) XML editing tools that would allow for permissions control, global change, etc. And ways to take advantage of our ILS authority records in our XML digital library records and keep the corresponding "bibliographic" data records in sync.
- 12) We desperately need a METS generation tool allowing a user to create a fileSec and structMap through a graphical interface. We'd like to start development on this tool soon, but we currently don't have a firm timeline. We're also going to start development soon on a MODS-based "generic" photograph cataloging interface with Oracle on the back end.
- 13) METS tools (edit, display, search) tools that transform metadata from one format to another tools to create application profiles - meta management of metadata! more tools like EAD cookbook but for other formats
- 14) Better normalization scripts and other tools.
- 15) We wish we had something similar to MARC-Edit that is geared toward single records. We wish we had a simple, modular tool for updating individual records (upgrading, editing, based on various metadata schemas) We wish we had a universal translator, or a scalable way of delivering forms of metadata that aren't our native formats. We envision something similar to Jeff Young's MARC XML switching schema, but more fully fleshed out.
- 16) Yes, a METS editor. We're building one, but it would be nice if the community could work on sharing those already developed.
- 17) I wish we have more scripts that transformed metadata from one standard to another. For example, I'm surprised there isn't a script to take metadata from a MARC record and put in the TEI header format (or vice versa). I'd really like to start using METS more so I wish there were tools to help build and validate METS Profiles. I wish there was a way to do MODS Formats too, like setting up templates for specific types of resources.
- 18) An intuitive XML editor for the creation of EAD.
- 19) Tools associated with emerging METS profiles
- 20) a nice, easy to use, mets packager would be nice

SECTION 4 OF 5: OPEN QUESTIONS AND COMMENTS

**7. This survey focuses on the immediate concerns of the Aquifer Metadata Working Group. We recognize that there are other issues regarding metadata for cultural heritage digital resources that this survey does not address. If you have additional questions or comments that you feel are relevant to this project, please bring them to our attention by mentioning them in the box below.
(9 responses, 16 skipped question)**

- 1) We see a need for Shared Best Practices, and believe the Aquifer Metadata Working Group is in good position to recommend/specify practices to foster interoperability / eventual federation of collections that are currently discrete.
- 2) How does one determine whether MODS is the best standard to use in cultural heritage applications. We need more metadata transformation tools.
- 3) I think we all need to pay attention to the standards and tools we use for the content of our metadata. For example, AACR2, CCO, DACS, ONIX & BISG metadata best practices. We also need to pay attention to the use of content standards to designate subjects or classify resources. For example LCSH, AAT, DDC.
- 4) PLEASE NOTE THAT THIS RESPONSE DOES NOT RELATE TO SIGNIFICANT CURRENT DIGITAL MATERIAL RELATING TO 'AMERICAN CULTURE AND LIFE'; INSTEAD IT RELATES TO DIGITAL AND DIGITIZE COLLECTIONS AT [institution] ([URL]) Two priority areas for us are - Digital Rights Management - Digital Signing of Objects.
- 5) We would find any training materials or other documentation on MODS to be extremely useful.
- 6) A comment on our MODS documentation: While we are willing to share such documentation as we have, it would probably be of limited usefulness to the Aquifer group. It is pretty specific to our [local database-driven tool] implementation, and maps the metadata from its [local database-specific] format to METS/MODS format.
- 7) Our use of MODS is based on MODS' ability to nest components needing some extent of description below the top level of description which works well with METS. The core element set that we've been using and probably will continue to use for the immediate future in our mass production environment is Dublin Core qualified based. We will need to continue to receive all kinds of metadata schemas from our data providers and plan to map what we receive to a core set that moves fairly well so far between our DCq and MODS mapping. If the Aquifer MD WG plans to write a profile for MODS for cultural heritage resources, I would urge that profile be based on so-called MODS lite rather than full MODS for I don't think it's realistic to require full MODS or full MODS with MADS.
- 8) Metadata Enhancement Services. There is a dire need for a set of modular tools that can enable catalogers and metadata specialists to quickly and easily remediate metadata records in various formats, including DC and MODS, both as individual records and through batch processing of collections of records. Although there has been some initial research in this area, there is no elegant solution to this problem to date. Creating such a tool in the near future is a major goal of the [project] at [institution]. We also feel that there is great need for a central, searchable registry of

all digitized books and sound recordings (possibly movies as well) that are available as open access publications on the Internet, and another registry of such items that have been digitized but are available only through proprietary systems. This would accomplish three goals: 1) alerting scholars/researchers about these materials and their locations; 2) helping libraries to expose their digital collections; and 3) avoiding the duplication of digital items. (We're aware of the Online Books Page from UPenn, as well as Digital Masters, Digital Books Index and several smaller systems, but they are not comprehensive.)

- 9) Why is METS listed as a possible descriptive metadata standard? It is not. Descriptive metadata according to other standards (MARC, MODS, DC, etc.) can be encapsulated in it or pointed to by it, but METS itself defines no descriptive metadata. FYI, we are not creating MODS natively, but we are mapping MARC and VRA to MODS for OAI purposes.