



OHIO DIGITAL PUBLIC LIBRARY OF AMERICA (DPLA) SURVEY REPORT

December 4, 2015

Prepared by Tom Clareson, LYRASIS and Liz Bishoff,
The Bishoff Group

Table of Contents

Introduction	2
Key Findings	2
Survey Demographics	3
Digital Collections	3
How many have digital collections?.....	3
Composition of digital collections.....	4
Copyright and Intellectual Property Issues	4
Digital Asset Management Systems.....	5
Metadata Issues.....	6
Number of metadata records	8
Harvesting metadata records	8
Collaboration.....	9
Additional Comments	9

Introduction

As part of the Ohio Digital Public Library of America (DPLA) planning grant, the project steering committee conducted a survey of Ohio's libraries, museums, historical societies, and archives to identify the current digital activities being undertaken by Ohio's cultural heritage organizations. The information gathered through the survey will aid in planning the Ohio DPLA initiative.

The survey was distributed to Ohio's cultural heritage organizations by state agencies and through professional association lists, and was available from October 1-26, 2015. A total of 219 responses were received.

This survey was funded in part through a federal Institute for Museum and Library Services (IMLS) Library Services and Technology Act (LSTA) grant, awarded by the State Library of Ohio to the Columbus Metropolitan Library.

Key Findings

- Who has digital collections? Of the 219 responses, 128 (58.99%) of the respondents reported that they had digital collections, while 89 did not. Of those that did not, 30 (33.7%) plan to create digital collections in the next three years. Based on this information, the initial three years of an Ohio DPLA project would focus on 128 organizations. An additional effort could focus on those organizations who plan to initiate digitization projects so that they are prepared to contribute to the Ohio DPLA program in future years.
- What content is available? Respondents have created or collected a range of formats of materials. Now that the Steering Committee has identified which organizations have digital collections, it will be important to define what subject areas or themes will be targeted for harvesting. Discussions with DPLA will help guide the collection development policy.
- What digital asset management systems are used? The Ohio DPLA service center or sub-hubs will need to harvest metadata from locally managed digital asset management systems. The survey found that of the respondents who had digital collections or planned to, 38 did not have a system while 12 did not know what system they had. The most commonly used system was CONTENTdm. Several of the organizations used multiple systems. Ohio DPLA needs to consider how to address the issue of how to support organizations that have digital content but don't have a digital asset management system.
- Who is ready to participate in Ohio DPLA? DPLA has recommended that Ohio identify organizations based on readiness to have their collections harvested; and that the institutions be organized into three groups. Based on the survey, the organizations that are likely to be most ready to have their collections harvest would be those who have implemented the OAI-PMH protocol, followed by those whose digital asset management system supports the OAI-PMH protocol but who haven't implemented it. The third group might be those organizations who have a digital asset management system and can provide data in some other manner. Ohio DPLA might consider a fourth group, those organizations with digital content but without a content management system. The survey identified 23 organizations that have implemented OAI-PMH, and 7 have systems that support OAI-PMH.
- How many metadata records are there? While 128 respondents reported having digital content, the majority of the collections were small, less than 10,000 metadata records. Of the 128, 35

organizations had between 10,001-999,000 metadata records that were publicly accessible; the majority of these organizations had systems that supported OAI-PMH.

- What is the status of access to digital resources and copyright? Key to successful participation in the DPLA is having the rights to provide access to the digital resources. The survey queried respondents regarding their level of confidence in determining copyright; most of the respondents felt very confident. However, fewer respondents actually recorded information regarding rights.

The survey collected samples of copyright statements from more than 40 respondents; as part of the effort to prepare to harmonize the metadata for DPLA contribution, the metadata working group should review the DPLA and Europeana paper, *Rightsstatements.org White Paper: Recommendations for Standardized International Rights Statements*, http://rightsstatements.org/files/151002recommendations_for_standardized_international_rights_statements.pdf. This paper provides guidance on the structure of rights statements in metadata records and recommends a dozen possible copyright statements that will ease understanding of rights for end users.

- Recommendation on the need for education: The survey identified a need for further education on issues related to digitization and born digital materials, copyright and other intellectual property issues such as deeds of gift, and metadata schemas, cataloging standards, and metadata harvesting practices. To build and sustain a successful Ohio Digital Public Library of America Service Hub, awareness and adoption of best practices and standards for digital creation and metadata development need to be spread across all types of cultural organizations.

Survey Demographics

Responses were received from a wide range of Ohio cultural heritage organizations, with public libraries having the highest response rate, with 107 (48.86%) responses. Sixty-eight (68) academic libraries responded (31.05%), and a combined group of archives (archives within a library, museum, historical society, or other type of organization) accounted for 29 or 13.2% of the responses. Historical societies and history museums, when combined, represented ten organizations or 4.56%.

A majority of the survey responses were from Administrators (Deans or Directors), with 135 responses (61.64%). Other titles/work roles of respondents included Librarian (35 or 15.98%), Digital librarian (9 or 4.11%), Archivist (8 or 3.65%), Information technologist (8 or 3.65%), Curator (5 or 2.28%) and Digital Archivist (3 or 1.37%).

Digital Collections

How many have digital collections?

When asked if their organization had digital collections, 128 respondents (58.99%) said yes, and 89 (41.01%) said no. Of those that currently did not have digital collections, 30 (33.71%) planned to begin creating digital collections within the next three years, and 59 (66.29%) did not. Thirty (30) of the organizations that do not plan to digitize collections were public libraries, and 21 were academic libraries. Those that did not have digital collections, and did not plan on creating digital collections in the next three years, were exited from the survey following this question.

Composition of digital collections

Those that are creating collections are digitizing a wide variety of material types. Among the most popular material types digitized by Ohio cultural heritage institutions (respondents could indicate all types of base materials digitized) were:

- Photographic prints: 108 respondents (80.6%)
- Text, manuscripts, and other multi-page items: 84 (62.69%) – especially popular among archives within a library
- Newspapers on paper: 67 (50%) – this category was very popular with academic and public libraries
- Books: 61 (45.52%)
- Maps, architectural drawings, posters, and other flat works on paper: 61 (45.52%) – all of the historical societies responding to the survey noted that they were digitizing these formats of material
- Microfilm/microfiche – any format: 45 (33.58%)
- Film – film negative or glass plate negatives: 45 (33.58%)
- Analog audio: 40 (29.85%)
- Two-dimensional works of art: 37 (27.61%)
- Analog video: 37 (27.61%)
- Journals on paper: 35 (26.12%)
- Three-dimensional artifacts: 29 (21.64%)
- Three-dimensional works of art: 23 (17.16%)
- Theses and dissertations: 23 (17.16%)
- Other material types: 15 (11.19%); eight of these respondents noted that they are digitizing yearbooks

Respondents have a range of born digital or already digitized materials including:

- Documents – PDFs, Word documents, spreadsheets, etc.: 99 (75.57%)
- Digital photography: 88 (67.18%)
- Digital video: 58 (44.27%)
- Digital audio: 54 (41.22%)
- Social media – blogs, websites, listservs, mailing lists, etc.: 45 (34.35%)
- Maps: 38 (29.01%)
- Electronic Theses and Dissertations: 28 (21.37%)
- Non-licensed e-books, e-journals: 21 (16.03%)
- Art or visual materials with database or digital component: 20 (15.27%)
- Research data, Geospatial data, Numeric data: 20 (15.27%)
- Software applications or operating system: 15 (11.45%)
- Other material types: 14 (10.69%)

Copyright and Intellectual Property Issues

Copyright and intellectual property issues can impact an organization's decisions regarding digital activities. Users of the Digital Public Library of America will link to the digital resource, and it will be necessary for the Ohio cultural heritage organization to have the right to provide access to the digital

resource. Respondents were asked to rate the extent to which a number of statements about copyright and other intellectual property issues were true at their organizations.

A majority of the respondents felt that the statement “we consider copyright and/or intellectual property concerns in managing digital collections” was very accurate (102 or 66.23%) or accurate (36 or 23.38%).

Most respondents “feel confident making copyright licensing and digital copyright decisions about (their) digital collections.” Thirty-three (21.43%) feel this is very accurate; 67 (43.51%) accurate.

While the above results show that Ohio organizations are confident in making decisions, they indicated that they have been deterred in digitizing collections due to copyright issues; “copyright and licensing concerns deter us from creating and preserving digital collections.” Forty-four respondents (28.76%) felt neutral about this statement; 33 (21.57%) felt it was somewhat inaccurate; 30 (19.61%) felt it was accurate.

There was a wide range of ratings for the statement “we record and maintain rights metadata to limit delivery of collections to authorized users.” While 58 respondents (38.67%) said this was not at all accurate, 32 (21.33%) were neutral, and 31 (20.67%) – including 13 academic libraries -- said this was accurate.

An even wider range of ratings was shown for the last statement in this section, “we have collections with restricted rights to disseminate or provide access.” Sixty-five respondents (42.76%) said this was not at all accurate, but 35 (23.03%) – including 18 academic libraries -- said it was very accurate.

Another question on rights asked if organizations have donor agreements that include rights to digital holdings. Sixty-two respondents (42.47%) said no, 50 (34.25%) said yes (20 of these were academic libraries), and 34 (23.29%) – including 17 public libraries -- said they did not know. Forty-eight respondents shared examples of the rights statements that are included in their metadata.

Digital Asset Management Systems

Ohio’s cultural heritage organizations reported utilizing a wide variety of digital asset management systems to manage their digital collections. The systems can be used to manage the full life cycle of digital objects, including metadata repository; image repository or linkage to the image repository; registry of preservation metadata; and a means of providing access to users. For the purpose of the survey, systems were listed that are also called Digital Repositories; Content Management Systems; Museum Management Systems; and Institutional Repositories. Respondents could choose all that apply and some are using multiple systems. It should be noted that 33 (25.98%) of the respondents do not have a digital asset management system; while 12 (8%) indicated they didn’t know.

- OCLC’s CONTENTdm (used by Ohio Memory, Cleveland Memory, Ohio Digitization hubs, etc.): 54 (36%)
- None: 38 (25.33%); 22 of these respondents were public libraries
- PastPerfect: 18 (12%), including all responding historical societies
- Omeka: 15 (10%)
- BePress Digital Commons: 15 (10%)
- Locally-developed system: 15 (10%)

- Internet Archive: 13 (8.67%)
- Don't know: 12 (8%)
- D-Space: 8 (5.33%)
- Fedora: 4 (2.67%)
- Hydra/Fedora: 4 (2.67%)
- The Gallery System: 2 (1.33%)
- Hathi Trust: 2 (1.33%)
- Islandora/Fedora: 1 (.67%)

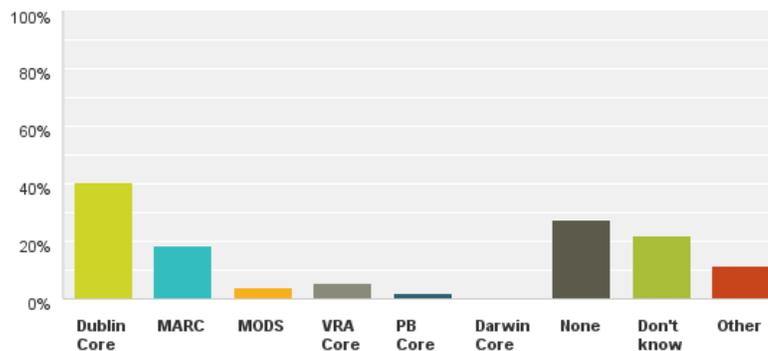
Twenty-four (24) respondents said they were using other systems; most of those were comments on past use of some of the systems listed above, consideration of adopting one of these systems, or single instances of use of systems which were not listed.

Metadata Issues

The survey asked which metadata schemas organizations are using for their digitization work.

Q14 What metadata schema are you using for your work?

Answered: 150 Skipped: 69



Dublin Core was used by 61 (40.67%) of the respondents; while 41 (27.33%) of the respondents said “none”, and 33 (22%), including 18 public libraries, said they did not know. Other schemas used by respondents included MARC (28 or 18.67%); Visual Resources Association Core – VRA Core (8 or 5.33%); Metadata Object Description Schema – MODS – (6 or 4%); Public Broadcasting Core – PB Core (3 or 2%). Seventeen (11.33%) responded “Other”; two of those were using Encoded Archival Description (EAD), and two were using METS-ALTO.

When asked about the cataloging standards and controlled vocabularies they are using in preparing descriptive metadata, with some respondents indicating multiple standards, the most popular choices were:

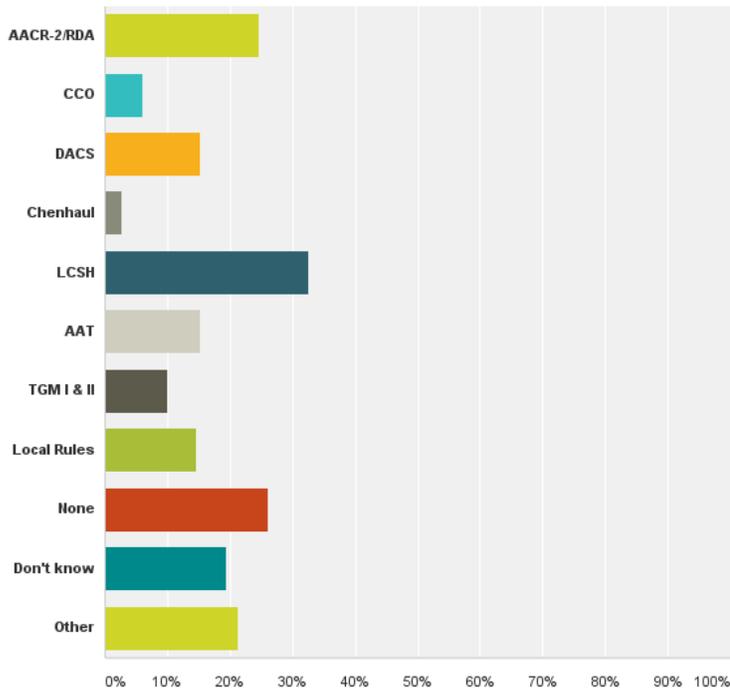
- Library of Congress Subject Headings (LCSH): 49 (32.67%), including a majority of archives within a library
- None: 39 (26%); 21 of these respondents were public libraries

- Anglo-American Cataloging Rules 2 (AACR-2)/Resource Description and Access (RDA): 37 (24.67%)
- Don't know: 29 (19.33%)
- Describing Archives: A Content Standard (DACS): 23 (15.33%)
- Local Rules: 22 (14.67%)
- Art and Architecture Thesaurus (AAT): 23 (15.33%)
- Thesaurus of Graphic Materials I and II (TGM I & II): 15 (10%)
- Cataloging Cultural Objects (CCO): 9 (6%)
- Chenhaul Nomenclature for Museum Cataloging: 4 (2.67%)

Twenty-six (26) organizations suggested other standards which mostly included local rules and subject headings.

Q15 What cataloging standards and controlled vocabularies are used in preparing descriptive metadata?

Answered: 150 Skipped: 69

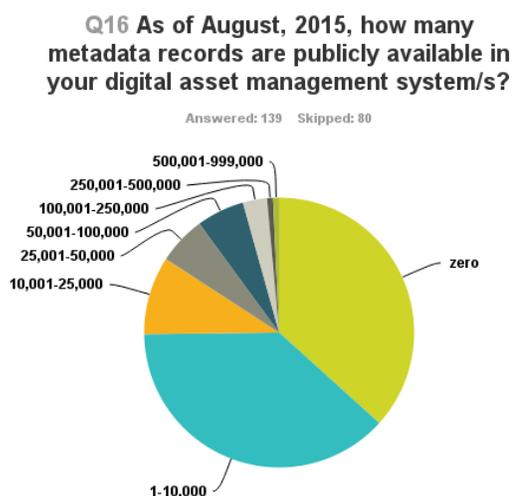


Number of metadata records

Organizations were asked, as of August, 2015, how many metadata records were publicly available in their digital asset management systems. This is important because these records can be harvested for inclusion in DPLA. The answer categories provided ranges of record numbers:

No of Records	No. of Organizations	Percentage
1-10,000	53	38.13%
Zero	51	36.69%
10,001-25,000	13	9.35%
25,001-50,000	8	5.76%
50,001-100,000	8	5.76%
100,001-250,000	4	2.88%
250,001-500,000	1	.72%
500,001-999,000	1	.72%

Eight organizations reported that they do not know the number of metadata records publicly available in their DAM systems.



Harvesting metadata records

A final question on metadata asked respondents what metadata harvesting strategy their organization supports. Metadata harvesting is the aggregation of metadata records from multiple providers into a single database.

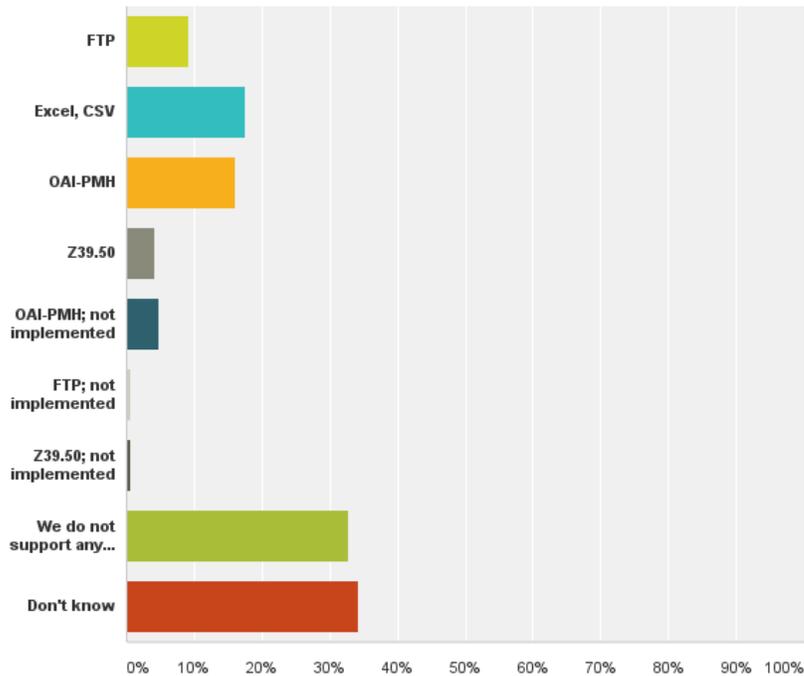
By far, the largest groups of respondents either did not know what capability they supported (49 or 34.27%) or do not support any metadata harvesting capability (47 or 32.87%). Among those that did support some type of harvesting, the most popular responses were:

- Excel, CSV (data export): 25 (17.48%)
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): 23 (16.08%)
- File Transfer Protocol (FTP): 13 (9.09%)
- Z39.50: 6 (4.2%)
- System supports OAI-PMH, but organization has not implemented it: 7 (4.9%)

- System supports FTP, but organization has not implemented it: 1 (.7%)
- System supports Z39.50, but organization has not implemented it: 1 (.7%)

Q17 What metadata harvesting strategy does your institution support?

Answered: 143 Skipped: 76



Collaboration

The final section of the survey asked if the respondents partnered with or belonged to an organization that has a digital collaborative initiative, such as Cleveland Memory. Seventy-eight (78) respondents (52.7%) said they do not participate in a collaborative digital initiative; 45 (30.41%) currently participate in such an initiative; and 14 (9.46%) participate in a collaborative initiative but are not currently active in it. Eleven (11) organizations (7.43%) did not know.

When asked which organizations they have partnered with, or what organization offers the digital collaborative initiative, a total of twenty collaborative groups were named; the most popular groups were Ohio Memory, Cleveland Memory, Columbus Memory/CML, the Digital Shoebox, Summit Memory, and the Five Colleges of Ohio.

Additional Comments

At the end of the survey, responding organizations were asked to share other thoughts that would inform the planning of the Ohio DPLA Project. Forty-one comments were received, most were from organizations that were interested in beginning digital projects, were very positive/supportive of the Ohio DPLA initiative, or had difficulty with the technical terminology in some of the survey questions.