Ohio DPLA Survey Report Draft 1: Interim Report to the Planning Committee

From Tom Clareson and Liz Bishoff

October 19, 20515

The Ohio Digital Public Library of America (DPLA) Planning Committee conducted a survey of Ohio’s libraries, museums, historical societies, and archives to determine the needs of institutions in planning for a statewide DPLA initiative.

**Survey Demographics**

The survey was sent to cultural heritage organizations through state agencies and professional association lists, and was available for response from October 1-16, 2015. By October 15, 199 responses had been received.

Public libraries had the highest response rate, with 103 (51.76%) responses. Fifty-nine (59) academic libraries responded (29.65%), and a combined group of archives (archives within a library, museum, historical society, or other type of institution) accounted for 23 or 11.5% of the responding organizations. Historical societies and history museums, when combined, represented seven institutions or 3.52%.

A majority of the survey responses were from Administrators (Deans or Directors), with 126 responses (63.32%). Other titles/work roles of respondents included Librarian (28 or 14.07%), Archivist (9 or 4.52%), Information technologist (8 or 4.02%, Digital librarian (6 or 3.02%), Curator (5 or 2.51%) and Digital Archivist (2 or 1.01%).

**Digital Collections**

When asked if their organization had digital collections, 116 respondents (59.18%) said yes, and 80 (40.82%) said no. Of those that currently did not have digital collections, 27 (34.62%) planned to begin creating digital collections within the next three years, and 51 (65.38%) did not. 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.

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: 94 institutions (78.99%)
* Text, manuscripts, and other multi-page items: 71 (59.66%)
* Newspapers on paper: 58 (48.74%)
* Books: 56 (47.06%)
* Maps, architectural drawings, posters, and other flat works on paper: 53 (44.54%)
* Microfilm/microfiche – any format: 39 (32.77%)
* Film – film negative or glass plate negatives: 36 (30.25%)
* Two-dimensional works of art: 32 (26.89%)
* Analog audio: 30 (25.21%)
* Analog video: 30 (25.21%)
* Journals on paper: 30 (25.21%)
* Three-dimensional artifacts: 24 (20.17%)
* Three-dimensional works of art: 20 (16.81%)
* Theses and dissertations: 18 (15.13%)
* Other material types: 12 (10.08%)

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

* Documents – PDFs, Word documents, spreadsheets, etc.: 90 (76.27%)
* Digital photography: 79 (66.95%)
* Digital video: 48 (40.68%)
* Digital audio: 43 (36.44%)
* Social media – blogs, websites, listservs, mailing lists, etc.: 42 (35.59%)
* Maps: 34 (28.81%)
* Electronic Theses and Dissertations: 24 (20.34%)
* Non-licensed e-books, e-journals: 17 (14.41%)
* Art or visual materials with database or digital component: 17 (14.41%)
* Research data, Geospatial data, Numeric data: 16 (13.56%)
* Software applications or operating system: 14 (11.86%)
* Other material types: 15 (12.71%)

**Copyright and other Intellectual Property Issues**

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

A majority of the respondents felt that the statement “we consider copyright and/or intellectual property concerns in managing digital collections” to be very accurate (92 or 67.15%) or accurate (30 or 21.9%).

Most respondents “feel confident making copyright licensing and digital copyright decisions about (their) digital collections.” Thirty-one (22.63%) feel this is very accurate; 59 (43.07%) accurate.

In an interesting finding, there were mixed responses to the comment “copyright and licensing concerns deter us from creating and preserving digital collections.” Thirty-nine respondents (28.86%) felt neutral about this statement; 32 (23.53%) felt it was somewhat inaccurate; and 26 (19.12%) 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 52 respondents (38.81%) said this was not at all accurate, 30 (22.39%) were neutral, and 24 (17.91%) 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.” Fifty-nine respondents (43.38%) said this was not at all accurate, but 30 (22.06%) said it was very accurate.

Another question on rights asked if institutions have donor agreements that include rights to digitize holdings. Fifty-eight institutions (44.27%) said no, 44 (33.59%) said yes, and 29 (22.14%) said they did not know. Thirty-seven organizations shared examples of the rights statements that are included in their metadata.

**Digital Asset Management Systems**

Ohio’s cultural heritage organizations reported a wide variety of answers to what digital asset management systems used to manage their digital collections. The systems can be used to manage the full life cycle of digital objects, including management of data creation; 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 25.98% of the institutions (33) have no digital asset management system.

* OCLC’s CONTENTdm (used by Ohio Memory, Cleveland Memory, Ohio Digitization hubs, etc.): 48 (37.8%)
* None: 33 (25.98%)
* PastPerfect: 15 (11.81%)
* Omeka: 12 (9.45%)
* BePress Digital Commons: 12 (9.45%)
* Locally-developed system: 11 (8.66%)
* Internet Archive: 10 (7.87%)
* Don’t know: 8 (6.3%)
* D-Space: 4 (3.15%)
* Fedora: 2 (1.57%)
* Hydra: 2 (1.57%)
* Islandora: 1 (.79%)
* The Gallery System: 1 (.79%)
* Hathi Trust: 1 (.79%)

Twenty-four organizations 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**

Organizations were asked about the metadata schemas they are using for their digitization work.

The schema used by most institutions was Dublin Core (52 or 40.94%). However, it is a concern that 34 (26.77%) responding institutions said “none” for the schemas used, and 26 (20.47%) said they did not know. Other popular schemas were MARC (23 or 18.11%); Metadata Object Description Schema – MODS – (6 or 4.72%); Visual Resources Association Core – VRA Core (4 or 3.15%); Public Broadcasting Core – PB Core (3 or 2.36%). Thirteen or 10.24% of those answering the question 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 institutions indicating multiple standards, the most popular choices were:

* Library of Congress Subject Headings (LCSH): 43 (33.86%)
* None: 33 (25.98%)
* Anglo-American Cataloging Rules 2 (AACR-2)/Resource Description and Access (RDA): 33 (25.98%)
* Don’t know: 22 (17.32%)
* Describing Archives: A Content Standard (DACS): 20 (15.75%)
* Local Rules: 20 (15.75%)
* Art and Architecture Thesaurus (AAT): 19 (14.96%)
* Thesaurus of Graphic Materials I and II (TGM I & II): 11 (8.66%)
* Cataloging Cultural Objects (CCO): 8 (6.3%)
* Chenhaul Nomenclature for Museum Cataloging: 4 (3.15%)

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

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:

* Zero: 43 (36.75%)
* 1-10,000 records: 43 (36.75%)
* 10,001-25,000: 12 (10.26%)
* 25,001-50,000: 7 (5.98%)
* 50,001-100,000: 6 (5.13%)
* 100,001-250,000: 4 (3.42%)
* 250,000-500,000: 1 (.85%)
* 500,000-999,000: 1 (.85%)

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

By far, the largest groups of respondents either do not support any metadata harvesting capability (43 or 35.25%) or did not know what capability they supported (40 or 32.79%). Among those that did support some type of harvesting, the most popular responses were:

* Excel, CSV (data export): 21 (17.21%)
* Open Archival Information for Metadata Harvesting (OAI-PMH): 18 (14.75%)
* File Transfer Protocol (FTP): 11 (9.02%)
* Z39.50: 6 (4.92%)
* System supports OAI-PMH, but organization has not implemented it: 6 (4.92%)
* System supports FTP, but organization has not implemented it: 1 (.82%)
* System supports Z39.50, but organization has not implemented it: 1 (.82%)

**Collaboration**

The final section of the survey asked if the responding institutions partnered with or belonged to an organization that has a digital collaborative initiative, such as Cleveland Memory. Seventy (70) respondents (55.12%) said they do not participate in a collaborative digital initiative; 40 (31.5%) currently do participate in such an initiative; and eight (6.3%) participate in a collaborative initiative but are not currently active in it. Nine (9) organizations (7.09%) did not know.

When asked who the organizations were who 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, and the Five Colleges of Ohio.

**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. Thirty-one comments were received, most were from institutions that were interested in beginning digital projects, were supportive of the Ohio DPLA initiative, or had difficulty with the technical terminology in some questions.