

Statewide & Regional Stepping Stones To The National Digital Platform

A New Collaborative Michigan Initiative

Michigan Statewide Survey on Digital Preservation Readiness

Report & Analysis

May 13, 2019

Authors

Stephen Bennett (Project Data Assistant)

Matt Schultz (Principal Investigator)

Executive Summary

The purpose of the [Statewide & Regional Stepping Stones to the National Digital Platform](#) project is to assess the status of various digital preservation programs in Michigan and to determine readiness and interest for engaging broader collaborations across the state and within the surrounding region to further each institution's digital preservation efforts. In order to accomplish this, a survey questionnaire was created and sent to approximately 1000 preservation institutions across the state. Consisting of roughly 40 questions touching on institutional demographics; scope of digital collections; and the use of standards, technologies, and policies, the questionnaire was administered to institutions online via Qualtrics throughout the Fall of 2018. The questionnaire received 155 responses and respondents' answers were analyzed for the ultimate purpose of gauging the feasibility of potential collaborations for a future statewide digital preservation network.

Introduction

Across the U.S. cultural memory institutions are working together to collectively address their digital preservation needs in more open, transparent, and sustainable ways, rather than persistently outsourcing this core part of their public mission to private commercial vendors. This evidences itself through impressive state/regional networks sprinkled all throughout the country such as the [Alabama Digital Preservation Network \(ADPNet\)](#), the [Historically Black Colleges & Universities Library Alliance \(HBCU LA\)](#), the [Indiana Digital Preservation Network \(InDiPres\)](#), the [Texas Digital Library \(TDL\)](#), the [California Digital Library \(CDL\)](#), as well as others that are just getting off the ground, including the [Digital Virginias](#) network.

Michigan has determined to be yet another leading state in the effort to advance digital preservation at manageable scales by partnering with nearby neighbors and across all sectors. The [Statewide & Regional Stepping Stones to the National Digital Platform](#) project is emerging to be the catalyst to lay a foundation for a thriving new community of practice and set of shared technical solutions. The project envisions a future where it will be easier for institutions to avoid single points of failure for their digital collection holdings, where it will be affordable to securely replicate and deposit collections off-site with trusted peers, and to foster an open community that institutions can turn to for sound advice and practical assistance in the daily work of curating their digital collections. As this new community/network launches in Michigan the Stepping Stones team will be sharing all of its startup resources with other states and regions so that they too can kick-start their own community initiatives.

This work represents yet a further realization of the ongoing work of developing our National Digital Infrastructure and Initiatives (formerly National Digital Platform). Many thanks to the Institute of Museum and Library Services for their support of this project and all those advancing the NDII.

Methods

This statewide survey was designed first and foremost to gather data from collecting institutions across Michigan to better understand the maturity of their digital preservation programs and practices. Answers were sought for things like institutional demographics; scope of digital collections; and the use of standards, technologies, and policies. The survey was administered via Qualtrics and consisted of roughly 40 questions. After consulting the [RepoData Project for Michigan](#) and considering the collective membership lists of the four main professional associations that assisted with distributing the survey over the Fall of 2018—the included the Midwest Collaborative for Library Services (MCLS), the Michigan Academic Library Association (MiALA), the Michigan Archival Association (MAA), and the Michigan Museums Association (MMA)—the Principal

Investigator (Matt Schultz) determined that the survey had the potential to reach well over 1000 institutions across the state.

To ease the process of gathering the data the number of open-ended response questions was kept to a minimum. Yes/No/Not sure options were used as much as possible and where multiple choices were sought the list of possible answers was made as comprehensive, up-to-date, and relevant as possible. “Other” responses were always invited. Question logic was used for several key conditional responses, and respondents were permitted to skip any question they so choose with the ability to move back and forth within the survey. Respondents could also return to the survey and pick up where they left off if they needed to pause and gather more data from colleagues within the institution. Multiple responses from the same institution were permitted acknowledging that some parent institutions have multiple sites of curation that service different types of digital collections. Efforts were taken to disambiguate multiple responses from single institutions where analysis seemed to warrant.

The survey received 155 total responses. We heard from over 70 public libraries, over 30 academic libraries, nearly 30 museums, approximately 20 different university archives, over one dozen historical societies and research centers, nearly a dozen rare & special collections libraries, and at least three galleries. Our response rate for the survey overall was roughly 15%--which the project team considers quite healthy given the potential pool of respondents (1000). The full set of survey questions is available with this report as an Appendix, along with the full anonymized set of respondent data as a spreadsheet.

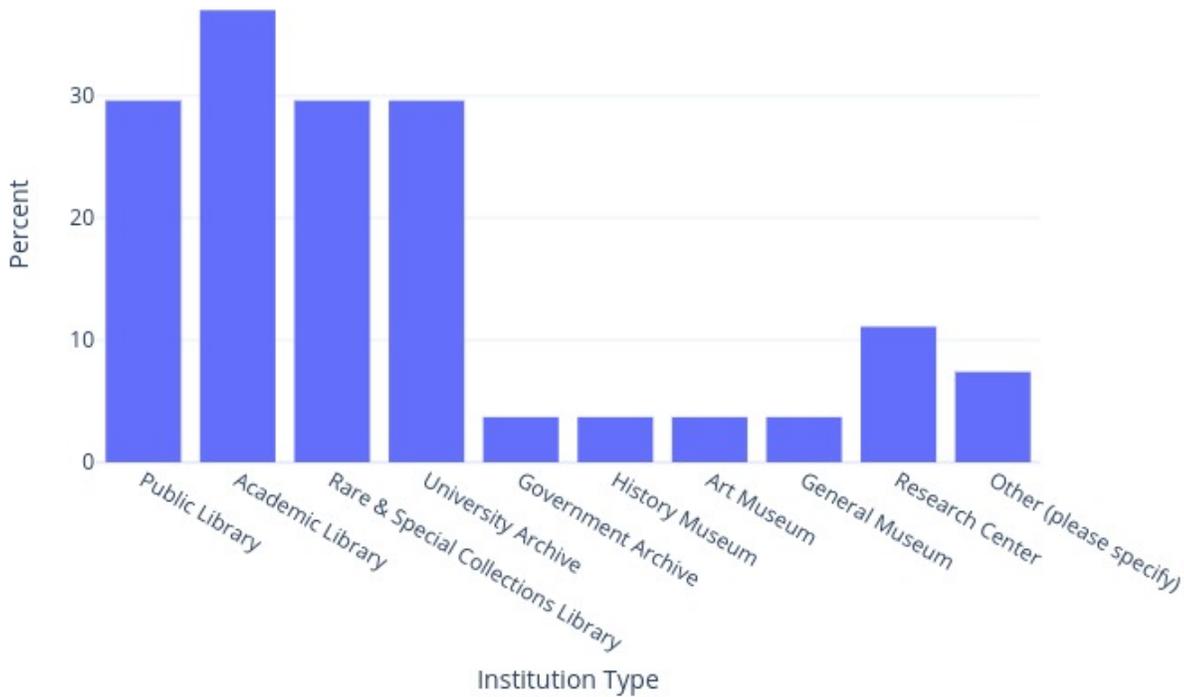
Correlations and relationships were sought between a number of questions, with the goal being to understand the state of digital preservation within the state in the context of considering future collaborations between institutions. In order to analyze the relationships between “check all that apply” questions, for example, we created new data tables for each of the relationships being analyzed. These data tables included variables from both questions, with the first set of responses being repeated for each category of the second set of responses. Counts and frequencies were then recorded based on each of these categories. We made sure to record both percent of boxes checked as well as percent of all cases when recording frequencies, although for our purposes we were more interested in percent of cases.

The analysis of the results that follows is only the start of forming the picture that is at target in the *Statewide & Regional Stepping Stones to the National Digital Platform* project. The project team is seeking clarity within the data to understand in what ways and along what lines a new network and community of practice might be able to get off the ground in the state of Michigan and for the surrounding region. It is looking for certain trends and seeking to understand barriers and challenges to collaboration—both organizationally and technologically. As we state in the Conclusion, the project team will continue to return to the data and will undoubtedly seek to gather more data in the later stages of the project to form a better picture. In the meantime, the project team invites any and all questions, critique, and feedback on this report. Please contact the Principal Investigator (Matt Schultz) at schultzm@gvsu.edu.

Report and Analysis

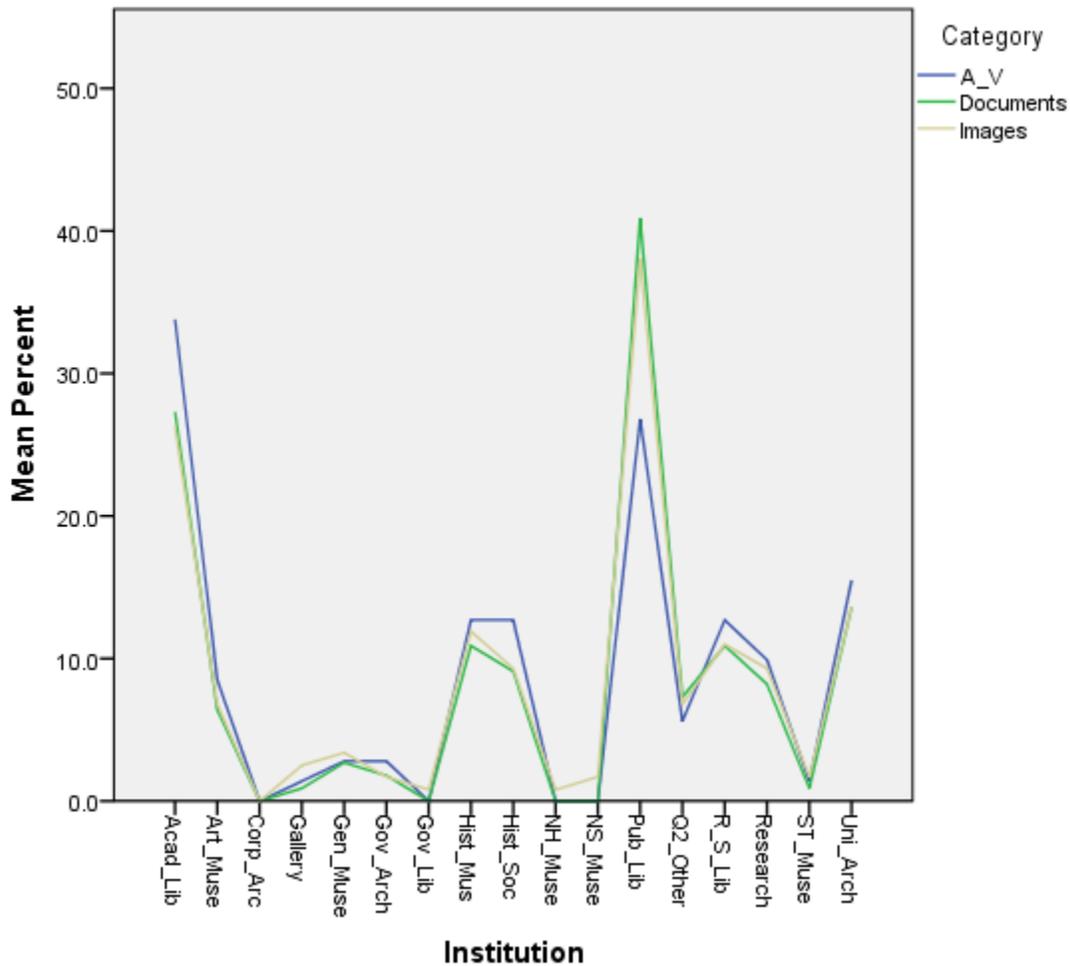
Data Amounts

Percent of Institutions Managing More Than 1 Terabyte of Data



The first thing we wanted to assess from our survey results was what types of institutions were holding the most data. The hope is that this could give us a baseline understanding of the institutions who might have the greatest need and whom would benefit most expeditiously from getting involved in a collaborative digital preservation network. Presumably the more data to steward the greater the perceived need for timely and transparent solutions. In order to answer this, we analyzed the frequency of the different institution types after isolating institutions managing one terabyte or more of data. Based on this analysis, we found that institutions identifying as academic libraries were the most frequent respondents of those managing more than one terabyte of data (37%). Public libraries are holding the most data averaging in the 500 gigabyte or less range. On average the museum sector reports managing data in the range of 500 gigabytes to one terabyte. A significant number of institutions reported not knowing how much data they were stewarding, which indicates a strong need for more inventorying at institutions across the state.

Data Types



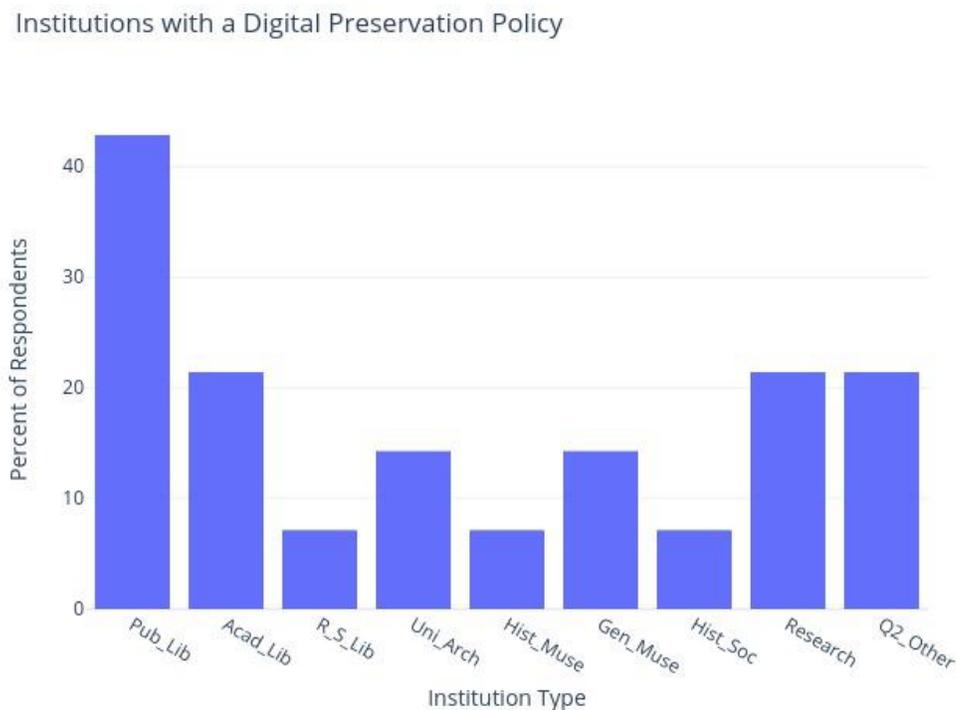
Next, we wanted to identify what the most prominent data types were that institutions were stewarding. This was done using a simple frequency command. We found that among our respondents, the top data types were images, documents, and audio-visual materials respectively. This is not incredibly surprising given the priorities of digitization for analog rare/special collections and the importance of oral history projects throughout the state. Databases and websites also figured somewhat prominently in importance amongst responses. This helps us to determine the major data types that a statewide network would need to be most poised to address in terms of content. After this we wanted to know if there was any significant relationship between institution type and popular data types. By percentage, the trends between different institution types are extremely similar. This is a good indication that a common solution aimed at addressing these core data types has potential for a good diverse contributor base of similar content—thereby laying a good foundation.

Data Complexity



Next, we were looking to examine the frequency (how many?) and relationship (overlap?) of those institutions holding digitized and born-digital content both separately and together. Knowing what this balance might be on institutional levels can provide a window into acquisition and production priorities, and indicate willingness to grapple with less complex to more complex digital materials. Here, we found that there appears to be a direct relationship between managing digitized content and managing born-digital content. When the respective responses for each question are isolated, the same response in the other question has the highest likelihood of being chosen. This makes it easier to draw conclusions for both groups. Thus, a common solution aimed at addressing digital content irrespective of its mode of creation/origin would seem to be entirely feasible and conducive to participation from a broad base of institutions and institution types. Experience with acquiring and managing both types of content is growing and increasingly commonplace.

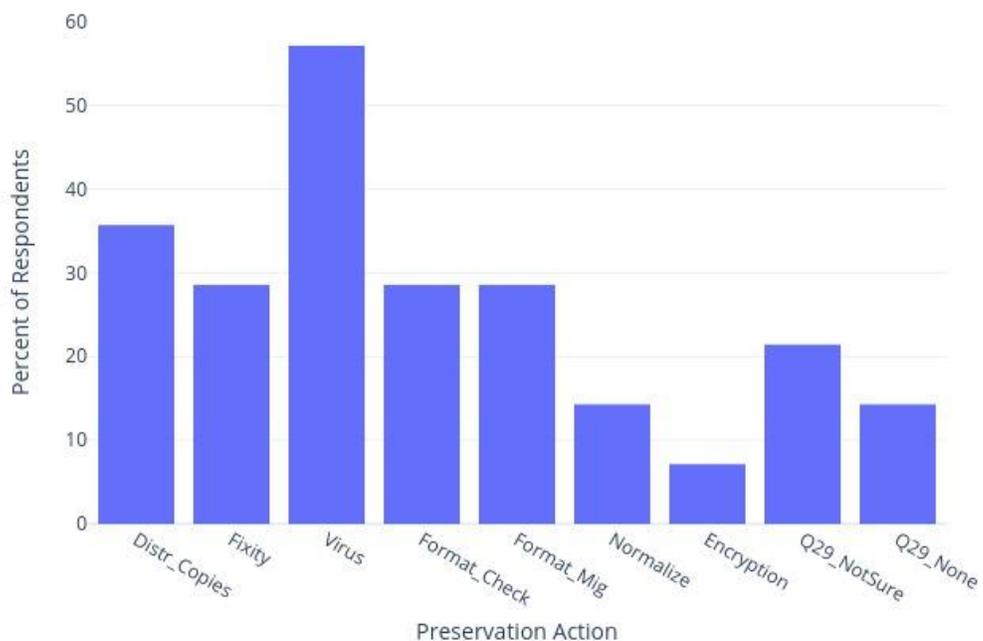
Digital Preservation Policies



Next, we looked at whether or not institutions had a digital preservation policy. Digital preservation policies can indicate a mature degree of thought and practice around responding to the challenge of digital preservation. Of some concern, we found that only 14 out of the entire 155 responses indicated that they had policies, and when directly solicited through the survey to provide links to those policies, unfortunately none were made available. A quick research investigation into the websites of those who indicated that they had policies indicated that in many cases what was most likely being cited were collection policies with very general references to preservation in the broadest sense for both analog and digital content.

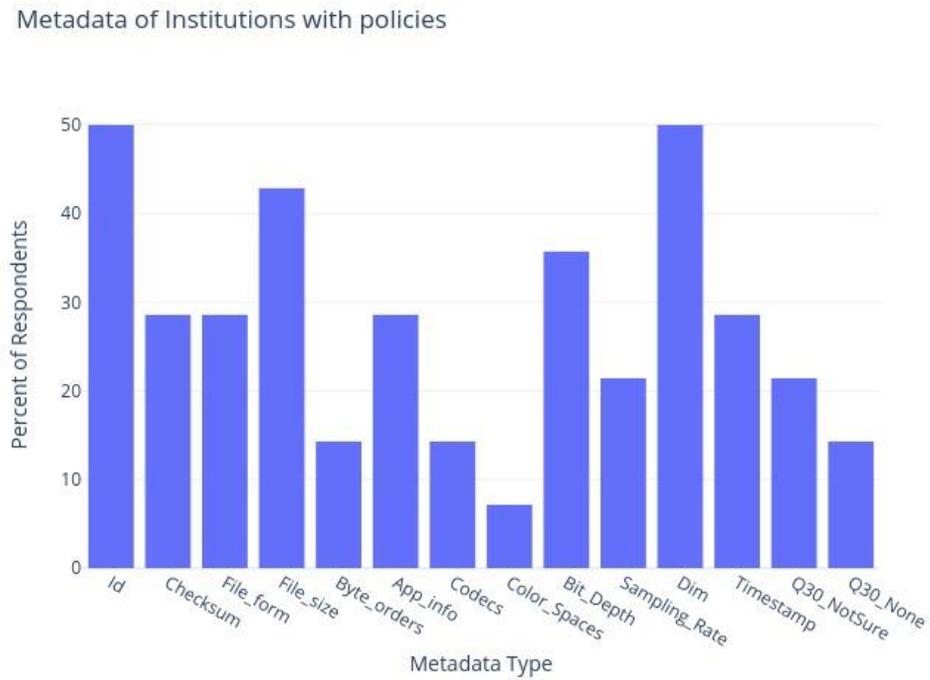
Preservation Standards & Practices

Preservation Actions from Institutions with Policies



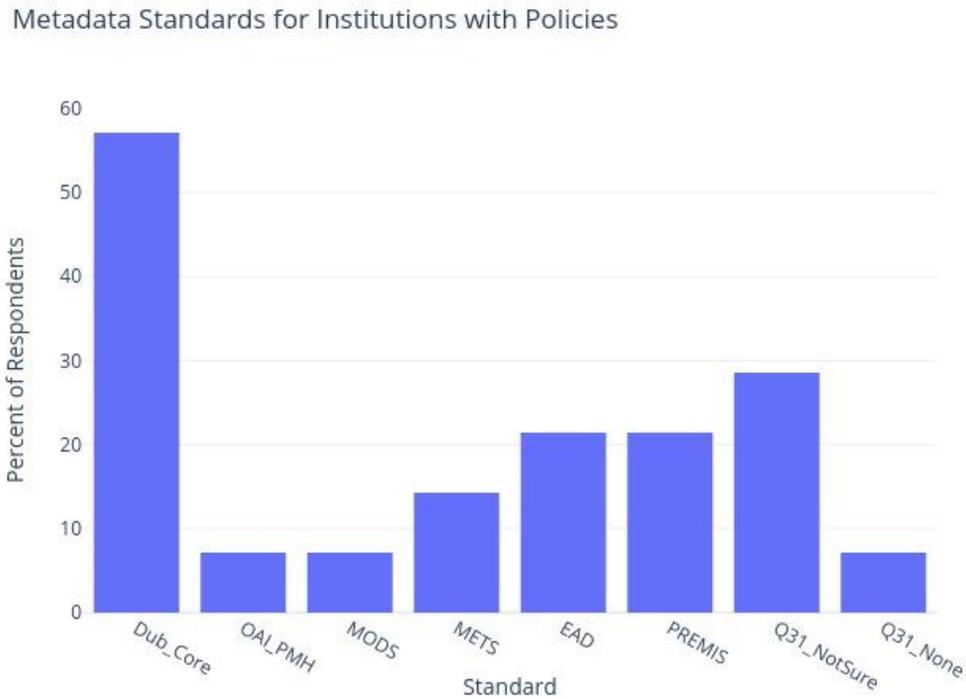
Of the institutions with policies, we found that virus checks were the most common policy-oriented digital preservation action cited, followed by geographic storage distribution, and then the performing of file fixity and format analysis. Very few institutions are performing more advanced approaches to digital preservation such as encryption. A not insignificant number of institutions reported None or Not Sure when it came to identifying their preservation actions.

Preservation Standards & Practices (cont.)



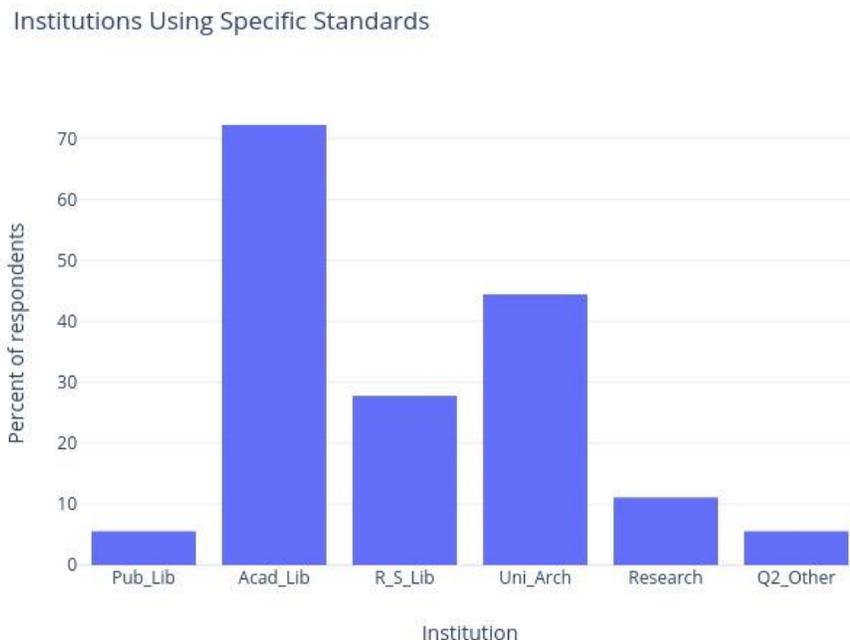
Lastly, we found that of the institutions with digital preservation policies, information such as image dimensions, image resolution, file sizes, and the assigning of unique identifiers were the most frequently occurring types of preservation metadata that these institutions are recording (more on this below).

Preservation Standards & Practices (cont.)



Dublin Core is far and away the most common type of metadata standard being used to contain the above-mentioned information and of course any descriptive information. The project team paid close attention to which institutions are using metadata standards like METS, PREMIS, and/or OAI-PMH. These are standards that are often used to help accomplish digital preservation, either through assisting with data transmissions or recording the outcomes of various digital preservation events. These standards can imply a fair degree of sophistication in the areas of technical infrastructure and skilled staffing.

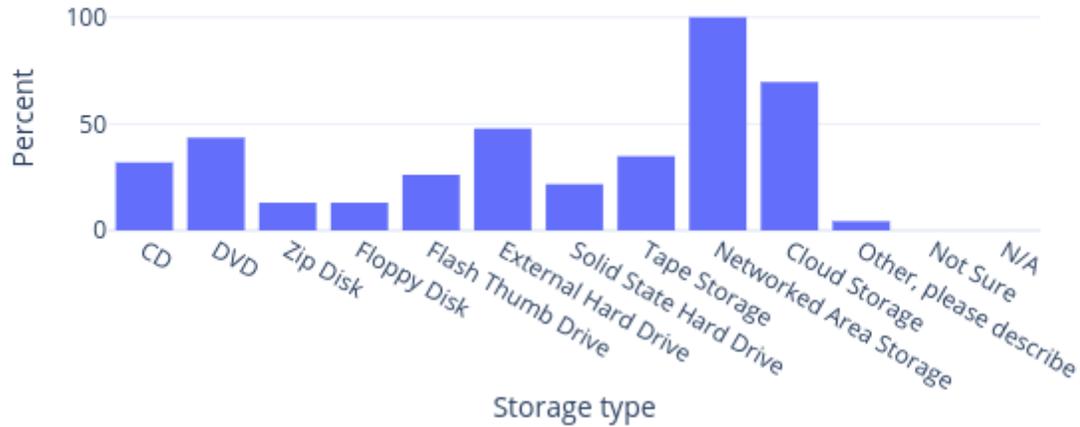
Preservation Standards & Practices (cont.)



On the whole there is not a huge degree of adoption and use of these preservation metadata standards across the pool of respondents. Not surprisingly, we found that academic libraries were far and above the most frequent institution type using these standards—they made up 72.2% of the respondents in the group. We also found that of the 17 different institution types listed, only a small subset was using these types of metadata standards at all: Public Libraries, Academic Libraries, Rare and Special Collections Libraries, University Archives, and Research Centers. This all would indicate to the project team that demanding interoperability with these standards more broadly to carry out data exchanges may not be fruitful or worthwhile—at least in the earliest stages. On the other hand, it is encouraging that a cluster of institutions across the state have some experience adopting, implementing, and maintaining these standards.

Optimal Storage Solutions

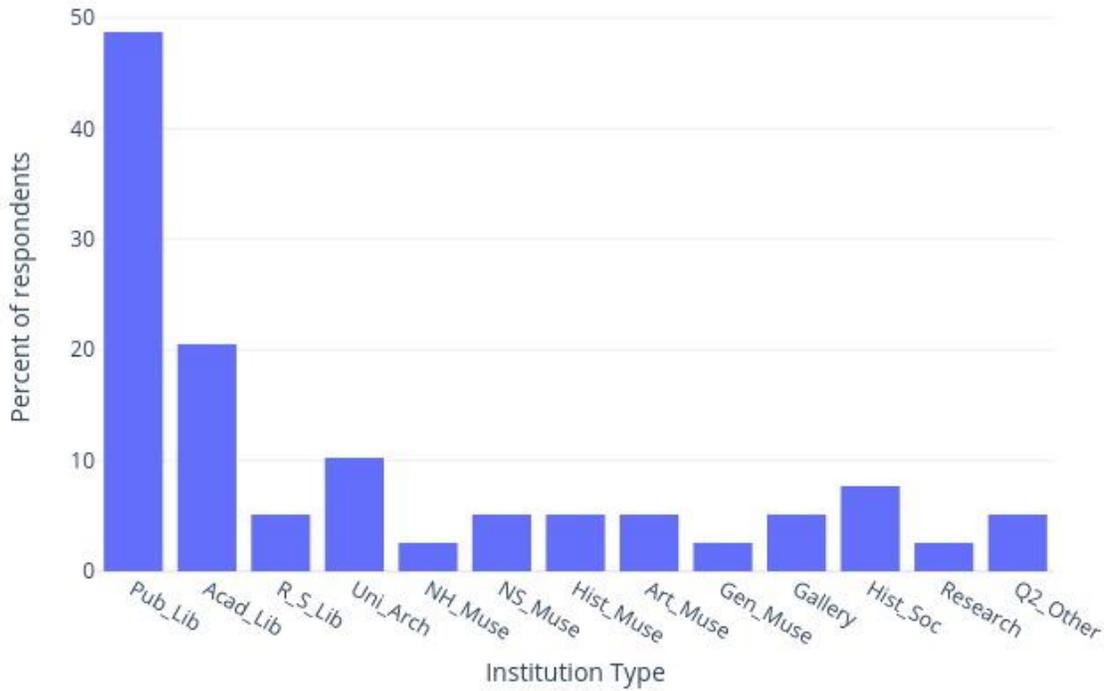
Storage types of high-storage institutions



We then sought to answer the question, “Of the institutions holding the most data, which storage types are most prominent. This would give our project team another metric for understanding how data might need to be interfaced with on a storage media and infrastructural level for the purposes of enabling cross-institutional data exchanges. Also, what might the standard level of expectation be for a future shared storage solution? For this case, we kept the same set of institutions managing more than one terabyte of data, and looked at the frequency of the different storage types within that group. By way of this investigation, we found that almost all of these institutions (92.6%) were using networked area storage in order to store their unique digital content. The next closest storage type was cloud storage, which 70.4% of the top storage institutions said they were using. Admittedly, external hard drives and increasingly risk-prone media such as CDs and DVDs were also being routinely relied upon, albeit to a lesser degree. This data gives the project team and partners starting points for considering functional requirements for a new network, as well as priorities for developing outreach, advocacy, and training resources to create pathways for longer-term stewardship of the content being held on the media.

Backup Practices

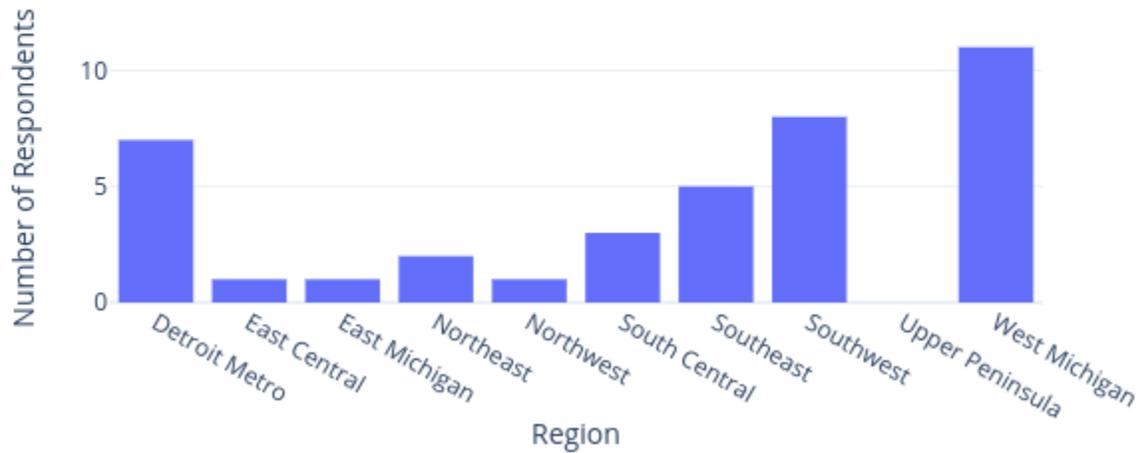
Institutions with No Backups



Next, we looked at geographic location of institutions who specifically did not have backup copies as part of their policy approaches to digital preservation. This indicates a tier of institutions at special risk for single points-of-failure. Getting digital collections out from under single points-of-failure has been a primary concern for this project. Nearly 40 institutions indicated that they did not perform backups on their digital collections. Almost half of those respondents were public libraries, followed by academic libraries, and university archives.

Backup Practices (cont.)

Respondents without backups by Region



Regionally, we found that the largest proportion of all the institutions (23.9%) were located in West Michigan, followed by Southwest Michigan, and Detroit Metro. It is also worth noting that there were no such institutions reported without backups located in the Upper Peninsula. Further follow-up and surveying on this latter finding are warranted.

Preservation Software

Technologies

- Content Management Systems



The next area of interest for us was to determine whether or not there was any sort of relationship between the institution types and the different software systems being used. This question proved difficult to answer as over half of the participants failed to respond to the question. Of those who did respond, many institutions had minimal representation, so an accurate relationship could not be identified. The vast majority of systems selected are geared towards facilitating discovery and access (e.g., PastPerfect, CONTENTdm, Digital Commons, ArchivesSpace, Omeka, Collective Access, etc.) as opposed to digital preservation.

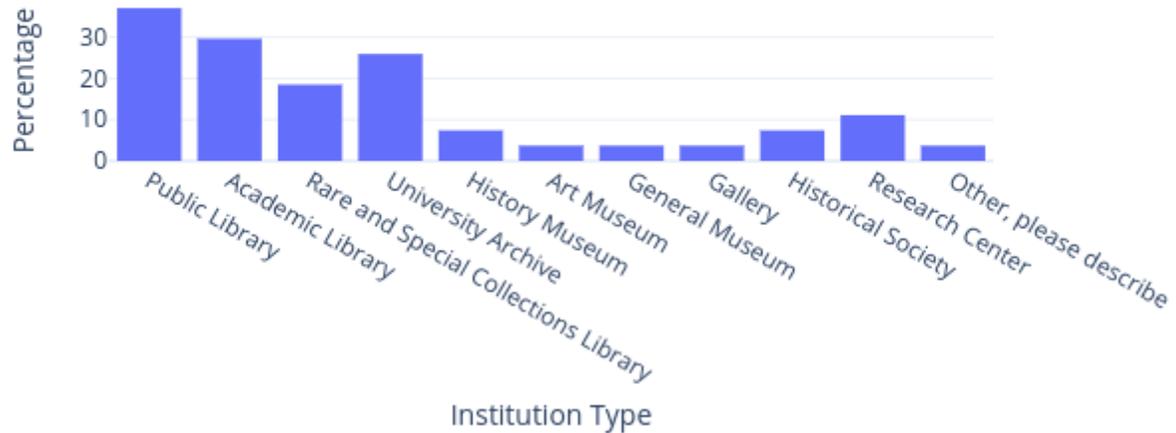
Technologies



Very few mature, all-purpose digital preservation systems (e.g., DuraSpace, Preservica, OCLC Digital Archive) are in use, and these primarily by academic libraries and rare/special collections libraries. Some academic libraries in the state are advancing some of the up-and-coming repository solutions that have been supported by initiatives like the National Digital Platform. These include open source solutions like Samvera and Islandora.

Data Portability

Institutions performing Network Uploads



After this, we wanted to examine how the different institution types were distributed among only institutions performing network uploads for the purposes of data transmission to their identified solutions—be they for access or preservation. For better or worse network uploads have been the prevailing standard for service-oriented digital preservation solutions for quite some time now, and it is anticipated as the primary protocol going forward. Which is not to say that other modes of transmission (for example shipped hard drives) can or should not be supported into the future—but these entail much more sophisticated offline human intervention and other time/distance latencies to account for. We found that among institutions performing network uploads, public libraries made up the highest proportion of institution types.

Institutional Breakdown of Mostly and/or Very unsatisfied (51):

- 18 (35.3%) public library
- 13 (25.5%) academic library
- 7 (13.7%) rare and special collections library
- 6 (11.8%) university archives
- 6 (11.8%) history museum
- 5 (9.8%) historical society
- 5 (9.8%) research center
- 3 (5.9%) art museum
- 2 (3.9%) government archives
- 2 (3.9%) natural science museum
- 2 (3.9%) general museum
- 2 (3.9%) other
- 1 (2%) natural history museum
- 1 (2%) gallery

Last but not least, in fact most importantly, we looked at which institutions responded that they were either “Mostly Unsatisfied” or “Very Unsatisfied” with their current digital preservation strategies. When taking this into account, we see that 32.9% of our respondents marked either “Mostly Unsatisfied” or “Very Unsatisfied”. From this group, 21.9% said that they were mostly unsatisfied and 11% said that they were very unsatisfied. In this 32.9%, the most prominent institution types are Public Libraries, Academic Libraries, and Rare and Special Collections Libraries. Of these institutions, 66.6% said that they would be interested in collaborating in a statewide digital preservation network.

Conclusions

Even though academic libraries continue to hold some of the largest amounts of data within the state on single institutional levels, it is important to note that public libraries are not far behind and were the largest pool of respondents to the survey. Academic libraries continue to lead the charge in the adoption of more robust standards, practices, and technologies. They also have access to the most networked storage resources and infrastructures to facilitate efficient data exchanges. However, public libraries are also not far behind in this regard. These two sectors share enough affinities and an education base in the area of digital preservation to perhaps consider exploring as a potential stable base of collaboration upon which to build broader community collaborations.

Looking across all of the respondents to the survey it is clear that the perception of which types of digital materials are in self-identified need of more digital preservation from institution to institution are quite common and clear. This is encouraging and could help speed-up early stages of collaboration around the preservation of shared digital content.

There is much more work to do within the state to advocate for standards adoption and more policy development, and with all of that, improved training on practices and the use of technologies. More inventorying of digital collections and the use of software is definitely still needed. A clear picture of the danger zones within the state has emerged and should help focus the efforts of any newly emergent digital preservation network/community.

More targeted follow up and research is likely needed to better understand the responses from museums, historical societies, and research centers in order to better clarify how in-line or perhaps divergent their needs and situations are from some of the clearer trends that emerged for Public Libraries, Academic Libraries, Rare and Special Collections Libraries, and University Archives.

On the whole, given the gaps and needs surfaced by the survey, and the fact that nearly 1/3 of all respondents communicated a high degree of dis-satisfaction with their current digital preservation strategies, the data would indicate that there is some significant and rewarding collaborative work to explore. This report recommends that those institutions that have achieved the greatest strides in resourcing their collections and infrastructures, and maturing their practices in line with policies, take the lead in forming a foundational base for collaboration and continue to work with this survey data to refine outreach and assessment to clarify the current findings.

The Statewide & Regional Stepping Stones Project will continue to foster this objective. Some of the first steps have already been taken through the Project Symposium held at Grand Valley State University on March 4th & 5th 2019, where the first summary of this data was shared with attendees, all of whom were respondents to the survey. The Symposium White Paper will be made available on the project website in June 2019. Additional community development activities are scheduled for the Summer and Fall and will be summarized in the paper.