Report to the Content Creation and Dissemination Team
Digital Preservation Working Group
June 17, 2016

Summary

The Digital Preservation working group, convened in January 2016, was charged with "developing digital preservation strategies." We realized, however, that in order to recommend such strategies, we needed to first know the state of digital preservation across the Orbis Cascade Alliance. So, in order to generate an environmental scan of Alliance members' digital preservation statuses and practices, we conducted an online survey, receiving feedback from 32 of 39 Alliance members. Our survey results indicate that while unique digital content abounds across the Alliance (with a current estimated total of 149 TB of unique digital content), digital preservation practices are lacking for many institutions in key areas. The survey data, combined with write-in responses to specific questions regarding digital preservation needs, forms the basis of our recommendations.

Recommendations

There is obvious interest in this topic, given the high response rate to this survey. As such, the Alliance should seriously consider delivering additional, high-level digital preservation education via webinars or in-person presentations, as well as pointing to and publishing digital preservation resources via the Alliance website.

In order to improve the Alliance's overall digital preservation score, two actions would be particularly welcome to address, respectively, Alliance institutions' lack of implementation of file fixity and & data integrity practices and consistent lack of storage space in diverse geographic locations. In order to address documented institutional needs, as reported in the survey, the Alliance should take several additional steps to improve coordination and communication.

Accordingly, the working group recommends that the Alliance:

1. Investigate the cost and feasibility of establishing a hosted or cloud-based digital archive for digital archival content from Alliance members. Making available storage in locations with different disaster threats for Alliance members would greatly improve the levels of digital preservation for many members.
    - Note: the need for investigating a shared archive was demonstrated in the analysis of NDSA levels, and echoed in the survey comments.
2. Provide in-person or web-delivered education and assistance with integrating the creation and checking of fixity information into workflows, including providing information on available tools, as well as when/how and where/how to store information.

    - Instruction could also be used on how to record and store transformation metadata, which was another consistent lack across the Alliance.
3. Create or point to digital preservation training kits online, and develop checklists/quick-start guides for Alliance institutions looking to improve their digital preservation practices.
    - Note: Introductory levels of digital preservation education are still needed for many institutions across the alliance. However, delivering webinars on these activities seems inefficient given the many resources available online.
    - Some specific requests:
        i. Create an online resource that documents which institutions have readers of obsolete file formats.
        ii. Provide lists of tools for various tasks and contacts at institutions using them on Alliance documentation site.
        iii. Provide a template for a digital preservation plan.
        iv. Provide list of digitization vendors and who has used them.
        v. Quick / High Level Digital Preservation FAQ - what it is/isn't and why it is important.

Introduction/Methods

The digital preservation working group represents an increasingly growing collection of digital preservation expertise in the Alliance, and we wanted in some sense simply to advertise through our survey that 1) this expertise exists and 2) that those that possess it are happy to lend it as needed. In order to deploy that expertise most effectively, however, we needed first to understand what digital preservation activities were most urgently needed. We decided to use the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation, to help us better chart and quantify areas of strength and weakness. We considered doing this ourselves through interviews with individuals at each institution, but quickly realized a survey could gather this information more efficiently.

The survey consisted of two parts. The first was written in order to gather an overview of information regarding current practices, size of archive, staffing, and policies/education. The second section was composed with the NDSA Levels specifically in mind. We finished the survey with two questions asking respondents to describe digital preservation areas for which they desire additional instruction or Alliance assistance.

Overview

Alliance members are active in creating and managing both born-digital and digitized content. Nearly all members report that they create and/or manage digitized photographs (97%), books and maps (84%), and audio or audiovisual materials (87%). Electronic theses and dissertations are also widely managed (84%).

Managing and storing unique digital content, however, does not mean one is actively doing the work of digital preservation. As we wrote in the introduction to the survey, “while activities such as backing up data and digitizing content may be part of an overarching digital preservation effort, they do not constitute “digital preservation” for the purposes of this survey**.”** In order to not to waste the time of those institutions who do not engage in digital preservation activity, we used an early question to define digital preservation,[1] after which we asked if, according to this definition, members were engaged in digital preservation practices. Of 32 respondents, twenty-five answered yes to doing “digital preservation,” and the other seven, after answering no, were diverted to the final survey questions regarding digital preservation needs.

Digital Archival Content across the Alliance

[1] We used this language in the survey: *Digital preservation is defined as “the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the very long term.”*

On the whole, the Libraries in the Orbis Cascade Alliance manage about 149TB of digital content for preservation purposes. 9 institutions manage 5 or more TB, 5 manage between 1 and 4.9 TB, and 10 manage less than 1TB, while one institution wasn't able to estimate.

Total estimated one-year past growth in digital content throughout the alliance equaled about 37TB, or about a 25% increase overall. An estimated 86% of the total growth in digital content has happened over the last 5 years. This indicates that the Alliance has had to mature rapidly on all fronts in the last 5 years, also while dealing with other technological challenges

Total estimated growth in the next year is 62TB, or a growth of 42%. In five years, institutions roughly expect their digital content to double in size (190%), and many indicate that it could be much more, or think it is "too hard to predict".

For the most part, the work of creating/digitizing and describing this content is conducted in house. The storage of this content is primarily done in-house as well, but it's important to note that nine institutions currently use some sort of cloud storage service for long-term archival storage; five members who are not currently using cloud storage expressed an interest, with two specifically mentioning Amazon S3 as a possible option; and another two respondents left the door open to future use of cloud storage.[2]

Preservation Policies and Training

The majority of respondents do not have a digital preservation policy or strategy in place, although ten respondents expressed an interest in working towards one. Eight respondents indicated that they do have either a digital preservation policy in place or a strategy when dealing with digital preservation issues. Of those without a policy or strategy, two are currently developing them.

In similar numbers, eight of the twenty-five respondents indicated that they offer, or have offered, digital preservation education, training, or outreach. Specific offerings include training through the sustainable heritage network, grant-related training, and webinar hosting. Five members would like to develop some sort of digital preservation educational offerings. Seventeen of the responding institutions do not offer digital preservation education.

Digital Preservation Staffing

Although there were twenty-five respondents to this question, only twenty-two have a measurable percentage of staff engaged in digital preservation work. Two respondents have staff doing some level of digital preservation related work, but it is as part of the overall daily workflow and difficult to assign a percentage to. The remaining respondents have dedicated staff doing a measurable level of digital preservation. Five member institutions have at least one

[2] Ten have no plans to move content to the cloud.

dedicated full-time digital preservation staff, with one institution indicating that they have three. The majority of respondents report having less than one full-time employee.

All respondents acknowledge that an increase in digital preservation staffing from current levels would be ideal for their institutions.

Alliance NDSA Levels of Digital Preservation

As noted above, we mapped participant responses to the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation, which is "a tiered set of recommendations on how organizations should begin to build or enhance their digital preservation activities."[3] The NDSA is a growing and highly respected consortium of (mostly memory) organizations concerned with digital preservation. Their "levels" have been used by other organizations already; of particular note, USGS's Fundamental Science Practices Advisory Committee Data Preservation Subcommittee developed a set of recommendations based on the original NDSA levels guidelines.

From survey responses, we estimated NDSA levels for each institution in each category, then calculated the average NDSA level per institution and for the Alliance as a whole. Although the NDSA Levels of Preservation include the idea of a progression, where requirements in lower levels are prerequisites for requirements in upper levels, we found that responses did not always neatly fit these steps and adjusted our estimates.[4]

Note that levels of preservation may vary within an institution, for example, between content in an IR and content in an Archives department. This survey was designed to gather an overview of the status of digital preservation across the Alliance, and does not attempt to determine the NDSA levels within an institution. We encourage each institution to contact us if they would like our subjective determination of their NDSA levels.

NDSA levels across the consortium

Overall, the average NDSA level for Alliance institutions was 1.37 (out of 4), with the mean being 1. This is a good starting point for future measurements, but more interesting insights as to future actions to improve digital preservation practices across the alliance can be drawn from looking at Alliance institutions' average levels in individual categories.

---

[3] National Digital Stewardship Alliance. The NDSA Levels of Digital Preservation: An Explanation and Uses. Phillips, Megan et al, 2013. Accessed 2016-06-14 at http://ndsa.org/documents/NDSA_Levels_Archiving_2013.pdf

[4] When determining an institution's NDSA level for a category, we attempted to balance a level where responses did not neatly fit into one level. For example, an institution might record administrative metadata (level 2) and descriptive metadata (level 3), but not technical metadata (level 3) nor transformative metadata (level 2), and does not inventory their content (level 1). They would be placed in level 2, based on missing some requirements of levels 1 and 2, but fulling some requirements from level 3.

NDSA category	Storage & Geographic Location	File Fixity & Data Integrity	Information Security	Metadata	File Formats
Alliance average level	1.17	0.54	1.29	1.79	2.04
Alliance median level	1	0	1	2	2

Alliance strengths
There are no NDSA categories where the Alliance is particularly strong; as a whole, we do not exceed level 2 in any category. The Metadata and File Formats categories are the two strongest areas, with Alliance averages of levels 1.79 and 2.04 respectively. It should be noted that, of the 23 institutions that indicated whether they recorded transformation metadata (a level 2 requirement), only 5 responded that they did. The average Alliance level is bolstered by the high rates of descriptive metadata creation (a level 3 requirement), at 100% of respondents.

Alliance weaknesses
File Fixity & Data Integrity is clearly the Alliance's weakest area, with only 7 out of 25 respondents above level 0. Storage & Geographic Location is another area with room for improvement, with an Alliance average level of 1.16.

NDSA levels by institution

While we are not publishing individual institutions ratings, we would like to note that those institutions with the largest archives are, for the most part, those that correspond to the highest average NDSA levels. We also found this to be the case based on other responses in our survey.