

First Aid Training for Those on the Front Lines: Digital Preservation Needs Survey Results 2012

Jody DeRidder

“The dilemma for the cultural heritage preservation community derives from the lag between immediate need and the long-term transformation of digital preservation expertise.”¹

INTRODUCTION

Every day history is being made and recorded in digital form. Every day, more and more digitally captured history disappears completely or becomes inaccessible due to obsolescence of hardware, software, and formats.² Although it has long been the focus of libraries and archives to retain, organize, and preserve information, these communities face a critical skills gap.³ Further, the typical library cannot support a true, trusted digital repository compliant with the Open Archival Information System (OAIS) framework.⁴ Until we have in place the infrastructure, expertise, and resources to distill critical information from the digital deluge and preserve it appropriately, what steps can those in the field take to help mitigate the loss of our cultural heritage?

The very “scale of the digital landscape makes it clear that preservation is a process of triage.”⁵ While educational systems across the country are scrambling to develop training programs to address the problem, it will be years, if ever, before every cultural heritage institution has at least one of these formally trained employees on staff. Librarians and archivists already in place are wondering what they can do in the meantime. Those on the front lines of this battlefield to save our cultural history need training. Surrounded by content under digitization, digital content coming into special collections and archives, assisting content creators in their research and scholarship, these archivists and librarians need to know what they can do to prevent more critical loss. Even if developing a preservation program is limited to ensuring the digital content survives long enough to be collected by some better-funded agency, capturing records in open standard interoperable technology neutral formats would help to ease later ingest of such content into a trusted digital repository.⁶ As Molinaro has pointed out, those in the field need “the knowledge and skills to ensure that their projects and programs are well conceived, feasible, and have a solid sustainability plan.”⁷ For those on the front lines, digital preservation education needs to be accessible, practical, and targeted to an audience that may have little technical expertise. Since “resources for preservation are meager in small and medium-sized heritage organizations,”⁸ such training needs to be free or as low-cost as possible.

Jody L. DeRidder (jlderidder@ua.edu) is Head of Digital Services at the University of Alabama Libraries, Tuscaloosa.

In an effort to address these needs, the Library of Congress established the Digital Preservation Outreach & Education (DPOE) train-the-trainer network.⁹ In six one-hour modules,¹⁰ this training provides a basic overview of the framework necessary to begin to develop a digital preservation program. The modules formed the basis for three well-attended ASERL webinars in February 2012.¹¹ Attendee feedback after the webinars indicated a deep need for practical, detailed instruction for those in the field. This article reports on the results of a follow-up survey to identify the topics and types of materials most important to webinar attendees and their institutions for digital preservation, in the fall of 2012.

APPROACH

The survey was open from October 2 until December 15, 2012. Invitations to participate were sent to the following discussion lists: Society of American Archivists (SAA) Archives & Archivists (A&A), SAA Preservation Section Discussion List, SAA Metadata and Digital Object Round Table Discussion List, digital-curation (Google group), Digital Library Federation (DLF-announce), and the Library of Congress Digital Preservation and Outreach (DPOE) general listserv. Each invitation clarified that respondents need not be Association of South Eastern Research Libraries (ASERL) members in order to attend the free webinars or to participate in the survey.

The survey consisted of three questions, the first to determine the sources of digital content most important for respondents' institutions to preserve, and the second to identify the topics of greatest concern to respondents themselves. For these two questions, respondents were asked to rate the options as:

- Extremely important
- Somewhat important
- Maybe of value
- Not important at all

The first two questions are as follows:

Please rate the following sources of digital content in terms of importance for preservation at your institution:

- Born-digital institutional records
- Born-digital special collections materials
- Digitized collections
- Digital scholarly content (institutional repository or grey literature)
- Digital research data
- Web content
- Other

Please rate the following topics in terms of importance to YOU, for inclusion in future training webinars:

-
- How to inventory content to be managed for preservation
 - Developing selection criteria, and setting the scope for what your institution commits to preserving
 - Selecting storage options and number of copies
 - Determining what metadata to capture and store
 - Methods of preservation metadata extraction, creation, and storage
 - Legal issues surrounding access, use, migration, and storage
 - Selecting file formats for archiving
 - Validating files and capturing checksums
 - Monitoring status of files and media
 - File conversion and migration issues
 - Business continuity planning
 - Security and disaster planning at multiple levels of scope
 - Self-assessment and external audits of your preservation implementation
 - Developing your institution's preservation policy and planning team
 - Planning for provision of access over time
 - Other

After each of these questions, respondents were provided a free text field in which to add additional entries related to the “Other” entry.

The last question on the survey asked respondents whether they are members of an ASERL institution, since ASERL is supporting this series of webinars.

RESULTS

Of the 182 respondents, 37 (20.7 percent) self-identified as ASERL members, 142 (79.3 percent) as non-ASERL members, and three skipped the question. All respondents answered the first two queries.

Sources of Digital Content

For the complete set of respondents, the top three types of material considered extremely important for preservation were born-digital special collections materials (65 percent, 117 respondents), born-digital institutional records (62.7 percent, 111 respondents), and digitized collections (61.2 percent, 109 respondents). Digital scholarly content, digital research data, and web content trailed in importance, rated extremely important by only 37 percent (64 respondents), 33.9 percent (59 respondents), and 30.6 percent (52 respondents) respectively. In clarification, one respondent listed “born-digital correspondence (e-mail),” another listed “state government digital archival records,” a third asked for instructions for use of “Kodak’s new Asset Protection Film for preservation of moving and still images,” and one specified that by “special collections” she meant “audiovisual.”

The concern for A/V materials was echoed by some of the 8 respondents suggesting other content as extremely important: “born-digital moving image preservation” (an ASERL respondent), “best practices for preservation of different audio and video formats” (also an ASERL respondent), “born digital photographs and video of college events,” and a request for an “audio digitization workshop.” Additional “other” entries were copyright pitfalls, data security, and “very practical steps that very small institutions can take to preserve their digital materials (e.g. how to check digital integrity, and how often, selection of storage media, and creation of a ‘dark archive’).”

One ASERL respondent indicated that she did not rate “born digital” institutional and special collections materials as extremely important for preservation only because her institution does not yet have a system set up for these, nor do they yet collect many born-digital special collections. She clarified that she does think this is extremely important despite the seeming lack of interest on the part of her institution.

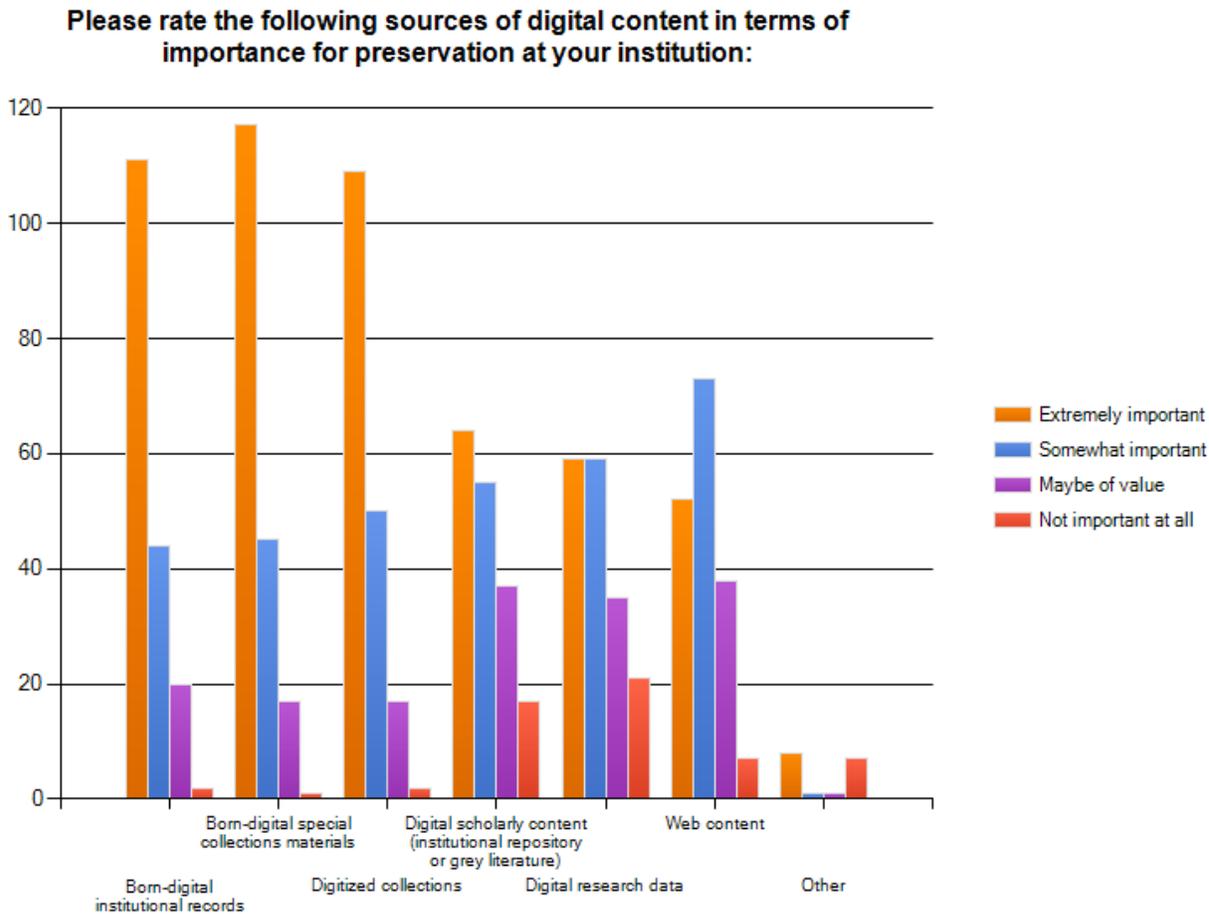


Figure 1. Results for all survey respondents indicating sources of digital content of importance for preservation at their institution.

In comparing the responses to the first question by whether the respondents self-identified as members of an ASERL institution (37 respondents as opposed to 142), those who did considered born-digital special collections materials far more important (73 percent, 27 respondents) than non-ASERL respondents (62.9 percent, 88 respondents), but this still was rated most important by both groups. Second for ASERL respondents was digitized collections (69.4 percent, 25 respondents) whereas born-digital institutional records held second place for non-ASERL respondents (62 percent, 85 respondents). Third and fourth-ranked material sources for ASERL respondents were born-digital institutional records (64.9 percent, 24 respondents) and digital scholarly content (63.9 percent, 23 respondents); digital research data only rated 52.8 percent (19 respondents). Non-ASERL respondents considered digitized collections the third most important source of digital content for preservation (59.7 percent, 83 respondents), and this group of respondents was far less concerned with digital scholarly content (29.9 percent, 40 respondents) or digital research data (29.6 percent, 40 respondents) than the ASERL respondents. Web content ranked lowest for both groups: 29.4 percent (10) ASERL respondents and 30.6 percent (41) non-ASERL respondents considered this content extremely important.

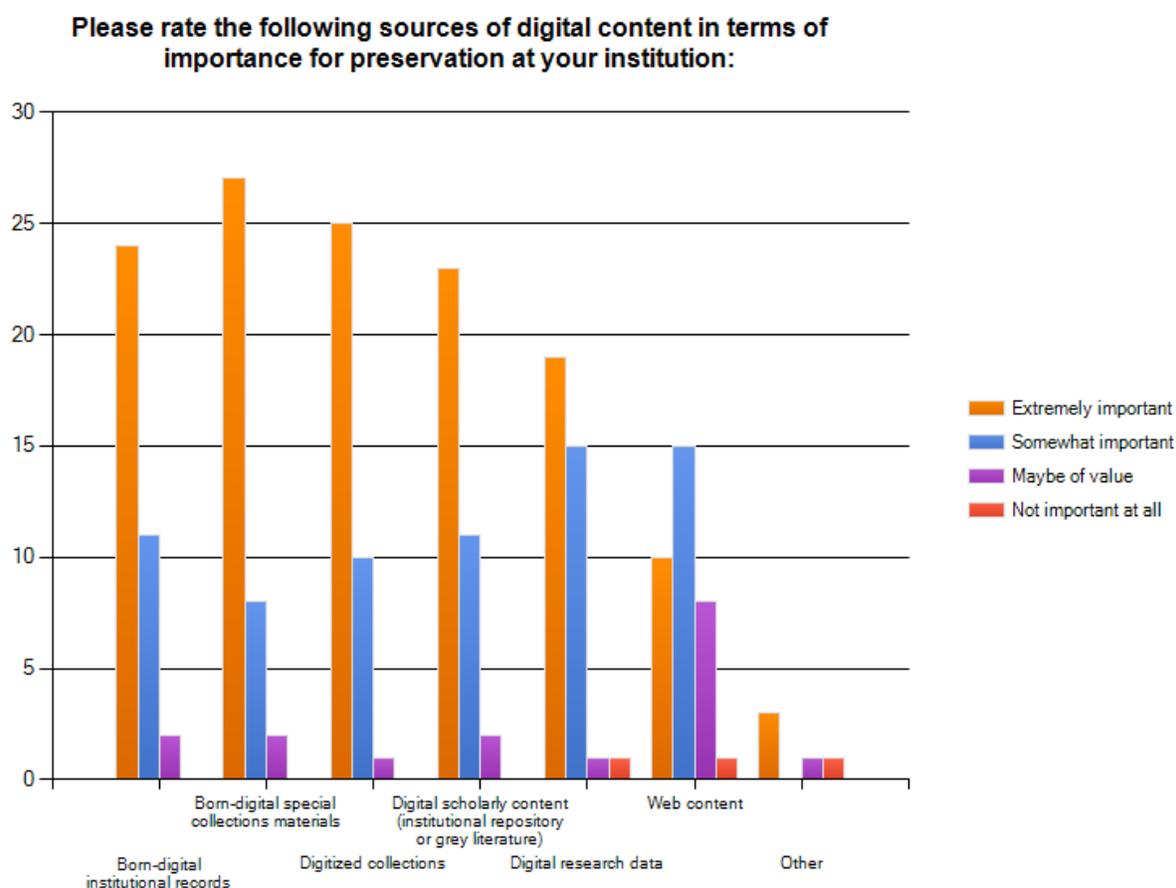


Figure 2. Results for ASERL survey respondents indicating sources of digital content of importance for preservation at their institution.

Please rate the following topics in terms of importance to YOU, for inclusion in future training webinars:

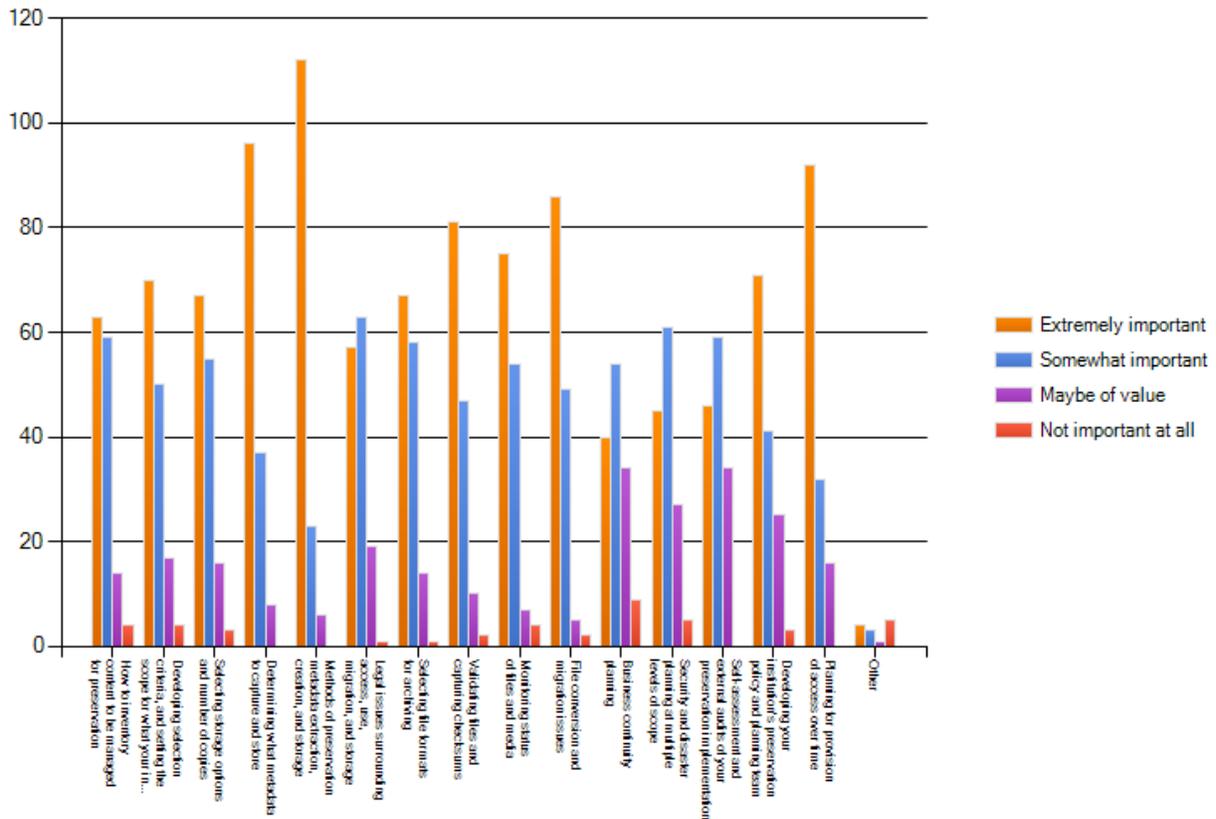


Figure 3. Results for non-ASERL survey respondents indicating sources of digital content of importance for preservation at their institution.

Perhaps most surprising was that 20 non-ASERL respondents (14.8 percent) rated digital research data as “not important at all” for preservation at their institutions, but this may be reflective of their type of institution. Museums and historical societies, non-research institutions, and government agencies likely are not concerned with research data; this theory seems to be supported by the 12.7 percent (17) non-ASERL respondents who rated digital scholarly content as “not important at all.” In comparison, only one ASERL respondent (2.8 percent) indicated that research data had no importance to his institution for preservation (0 for digital scholarly content). This may simply reflect a lack of awareness of current issues on the part of the respondent.

Topics of Interest

Both groups of respondents agreed on the three most important topics for future training webinars. “Methods of preservation metadata extraction, creation and storage” led the way with 77.3 percent (140 respondents: 70.3 percent or 26 ASERL and 79.4 percent or 112 non-ASERL)

listing this as extremely important. Next was “Determining what metadata to capture and store” (68 percent, 96 respondents: 62.2 percent or 23 ASERL and 66.7 percent or 120 non-ASERL). The third most important topic is “Planning for provision of access over time” at 65.4 percent (117 respondents: 1.1 percent or 22 ASERL and 65.7 percent or 92 non-ASERL).

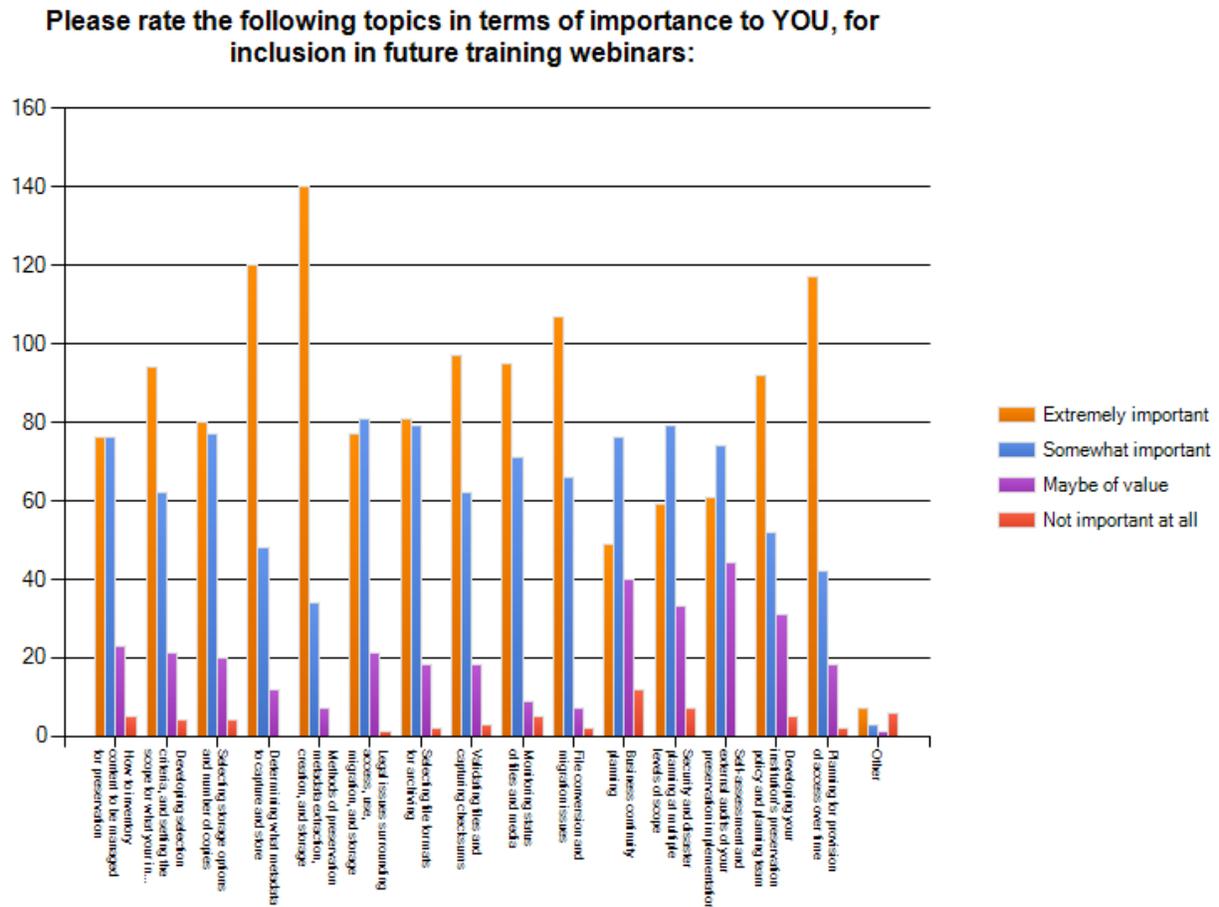


Figure 4. Results for all survey respondents indicating topics of importance to them, for future training webinars.

Fourth in importance overall was “file conversion and migration issues” (58.8 percent, 107 respondents: 54.1 percent or 20 ASERL and 60.6 percent or 86 non-ASERL), though the ASERL respondents thought this topic was slightly less critical than “developing selection criteria, and setting the scope for what your institution commits to preserving” (56.8 percent, 21 respondents as opposed to 49.6 percent or 70 non-ASERL respondents; overall percentage 51.9 percent, 94 respondents).

Close in relative importance were “validating files and capturing checksums” (53.9 percent, 97 respondents), “monitoring status of files and media” (52.8 percent, 95 respondents), and “developing your institution’s preservation policy and planning team” (51.1 percent, 92

respondents). Interestingly, however, “validating files and capturing checksums” is far more important to non-ASERL respondents (53.6 percent, 75 respondents) than those from ASERL institutions (only 37.8 percent, 14 respondents). “Legal issues surrounding access, use, migration and storage” is a more important topic for ASERL respondents (51.4 percent, 19 respondents) than non-ASERL (42.8 percent, 77 respondents), and ASERL respondents were more concerned (37.8 percent, 14 respondents) than non-ASERL (33.1 percent, 46 respondents) with “Self-assessment and external audits.” Additionally, “Selecting file formats for archiving” and “Selecting storage options and number of copies” is more important for non-ASERL (47.5 percent, 67 respondents and 47.9 percent, 67 respondents) than ASERL respondents (35.1 percent, 13 respondents and 32.4 percent, 12 respondents, respectively).

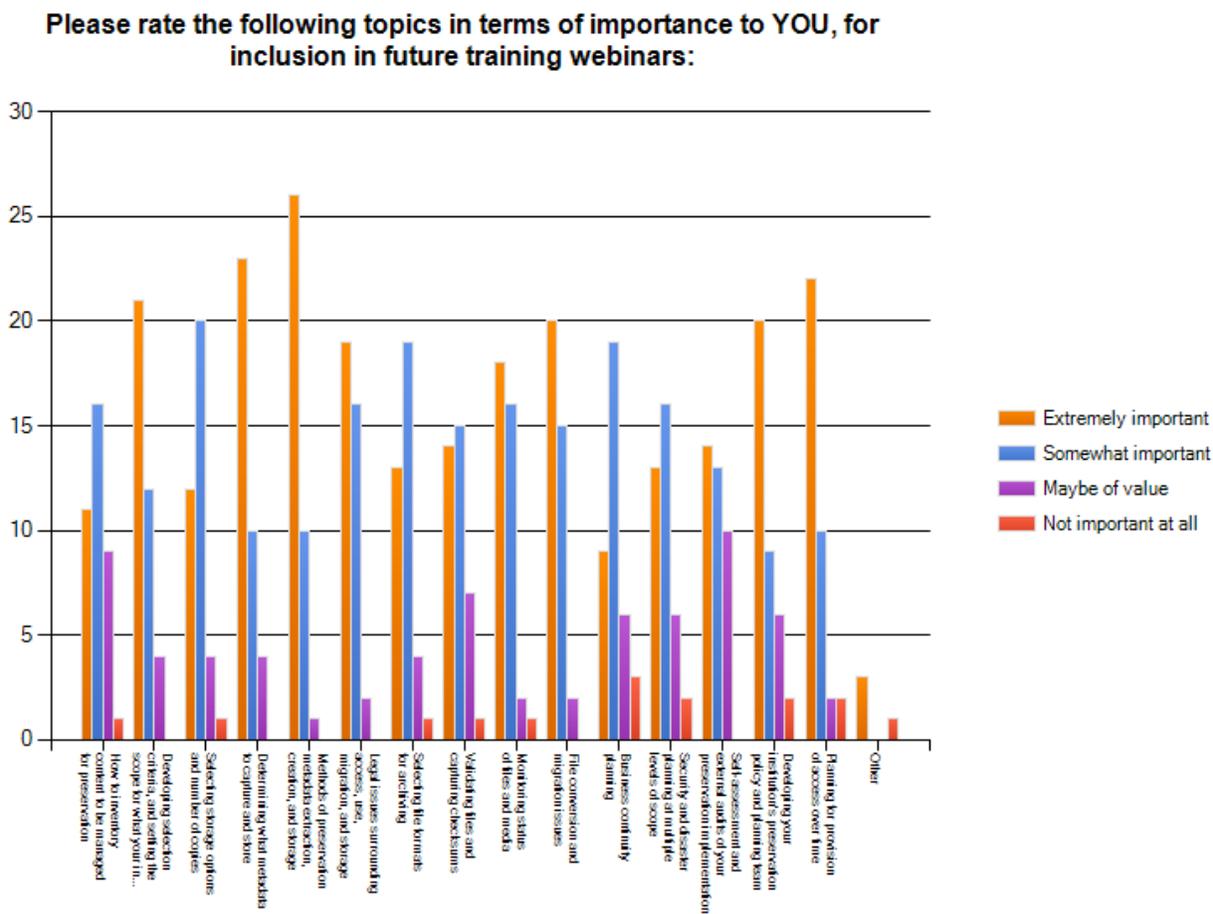


Figure 5. Results for ASERL survey respondents indicating topics of importance to them, for future training webinars.

“Security and disaster planning” was ranked extremely important by only 32.6 percent (45) respondents overall, followed by “Business continuity planning” at only 29.2 percent (40) respondents. The latter may reflect a lack of widespread awareness of just how critical the loss of

a single key employee can be, especially in smaller institutions. It also seems clear that there's a level of complacency or sense of security about our ephemeral digital content that may be in error. Then again, it is quite possible that the respondents are not administrators and feel they do not have the power in their organizations to address such issues.

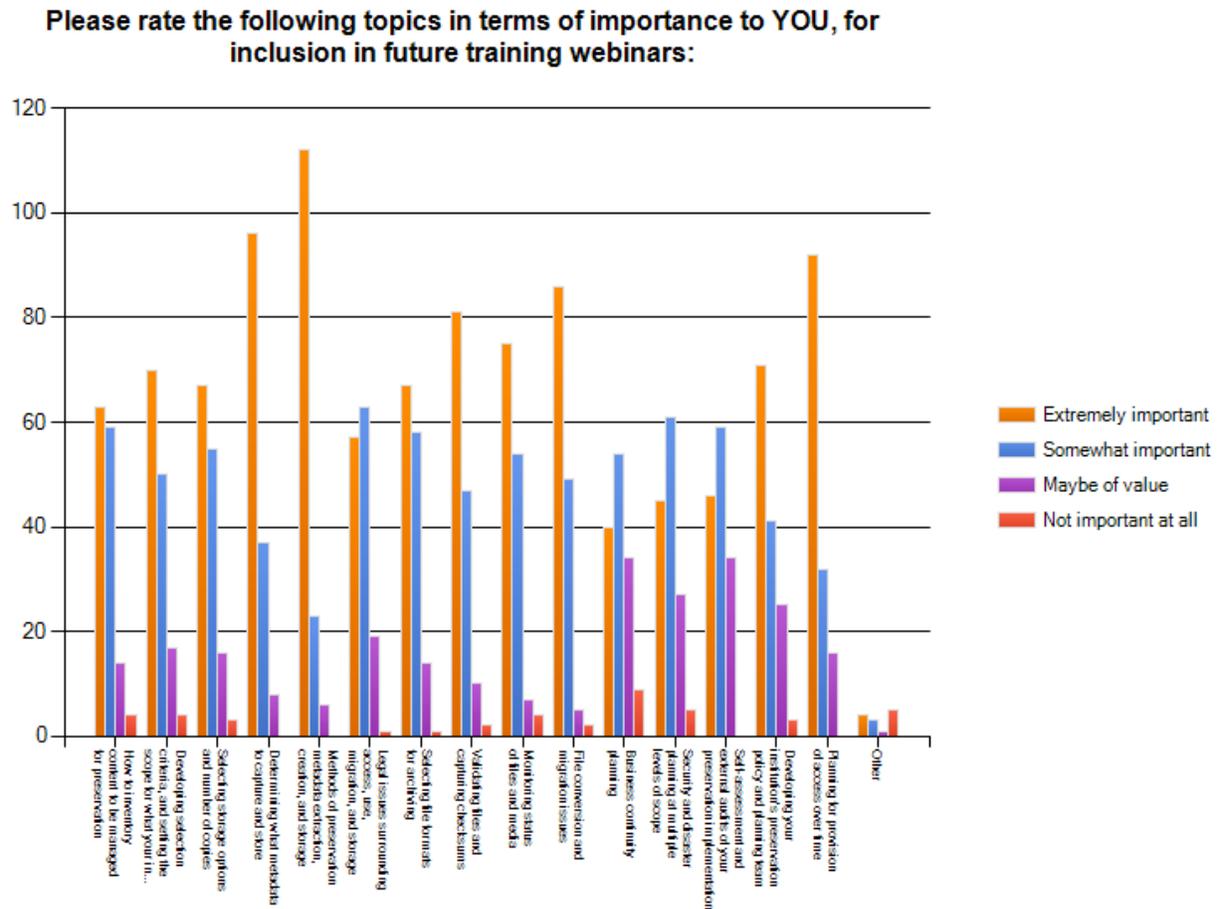


Figure 6. Results for non-ASERL survey respondents indicating topics of importance to them, for future training webinars.

Additional topics considered extremely important to respondents are as follows, listed in the free text area (the last four by ASERL members):

- "Clean" work station setup—hardware & software for ingest, virus scan, checksum, disk image, metadata, conversion, etc.
- Integrating tools into your workflow. There is a need to address the nuts and bolts for those of us that are further along in determining the metadata required to capture, selection criteria, and asset audit and preservation policy.

-
- Methods for providing researchers access to born digital content (not necessarily online, could be just in-house).
 - Strategies for locating digital assets on physical media in large collections that have been using MPLP [“More Product, Less Process”] for decades.
 - Format determination and successful migration or emulation.
 - Staff diversity and training.
 - How to validate files, migrate files, and which born-digital institutional files our special collections needs to be preserving.
 - Creating and maintaining effective organizational models for digital preservation (i.e. collaboration with Central IT and/or external vendors, etc.).
 - Case studies of digital preservation, establishing workflow of digital preservation.
 - Web archiving (best practices, alternatives to Archive-It, methods of selection, etc.).
 - One (non-ASERL) respondent said it was “somewhat important” to include the topic of “trends for field, future outlook.”

CONCLUSIONS

The results from this survey are clear: free or low-cost training needs to focus immediately on preservation of born-digital special collections materials, born-digital institutional records, and digitized collections. The topics of prime importance to respondents were “Methods of preservation metadata extraction, creation and storage,” “Determining what metadata to capture and store,” and “Planning for provision of access over time.” The variations in ratings between respondents from self-identifying as ASERL members versus non-ASERL members indicates that the needs of those in research libraries differs somewhat from that of cultural heritage institutions in the field dealing with “the long tail” of digital content.¹² Future training may need to target these differing audiences appropriately to ensure these needs are met. Additionally, administrators need to be addressed as a unique audience in order to focus on the requirements for addressing “Security and disaster planning” and “Business continuity planning,” as these critical areas need to be developed by those in management positions.

Future surveys of this nature should include a component to determine the level of technical expertise and support the respondents have, as well as a measure of their position or power in the administrative hierarchy. Continued surveys would be extremely helpful in ensuring that available educational options meet the needs of librarians and archivists in the field.

As Molinaro has pointed out, “Getting the right information in the right hands at the right time is a problem that has plagued the library community for decades.”¹³ Now is the time to develop free, openly available, practical digital preservation training for those on the front lines, if we are to retain critical cultural heritage materials which are only available in digital form. For them to effectively perform necessary triage on incoming digital content, they must be trained in “first aid.” Our history is at stake.

REFERENCES

1. Paul Conway, "Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas," *Library Quarterly* 80, no. 1 (January 2010): 73–74, doi:10.1086/64846.3.
2. Clifford Lynch, "Challenges and Opportunities for Digital Stewardship in the Era of Hope and Crisis" (keynote speech, IS&T Archiving 2009 Conference, Arlington, Virginia, May 2009).
3. Karen F. Gracy and Miriam B. Kahn, "Preservation in the Digital Age," American Library Association, *Library Resources and Technical Services* 56, no. 1 (2012): 30.
4. Marshall Breeding, "From Disaster Recovery to Digital Preservation," *Computers In Libraries* 32, no. 4 (2012): 25.
5. Mike Kastellec, "Practical Limits to the Scope of Digital Preservation," *Information Technology & Libraries* 31, no. 2 (2012): 70, doi:10.6017/ital.v31i2.2167.
6. Charles Dollar and Lori Ashley, "Digital Preservation Capability Maturity Model," Ver. 2.4, (November 2012), <https://docs.google.com/file/d/0BwbqtwrvKHokRXNVNmhXTmo2SUU/edit?pli=1> (accessed Dec. 24, 2012).
7. Mary Molinaro, "How Do You Know What You Don't Know? Digital Preservation Education," *Information Standards Quarterly* 22, no. 2 (2010): 45.
8. Conway, "Preservation in the Age of Google," 70.
9. Library of Congress, "Digital Preservation Outreach & Education: DPOE Background," accessed December 31, 2012, www.digitalpreservation.gov/education/background.html.
10. Library of Congress, "Digital Preservation Outreach & Education: DPOE Curriculum," accessed December 31, 2012, www.digitalpreservation.gov/education/curriculum.html.
11. Jody L. DeRidder, "Introduction to Digital Preservation—A Three-Part Series Based on the Digital Preservation, Outreach and Education (DPOE) Model," Association of Southeastern Research Libraries, 2012, [archived webinars], accessed December 31, 2012, www.aserl.org/archive.
12. Jody L. DeRidder, "Benign Neglect: Developing Life Rafts for Digital Content," *Information Technology & Libraries* 30:2 (June 2011): 71–74.
13. Molinaro, "How Do You Know What You Don't Know?" 47.