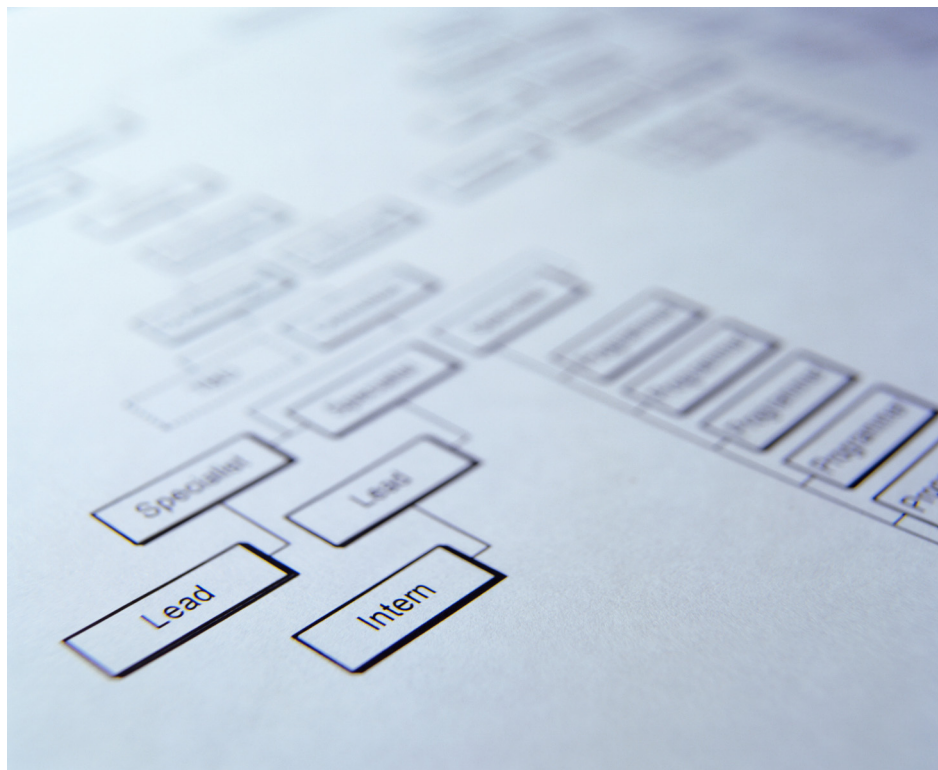


STAFFING FOR EFFECTIVE DIGITAL PRESERVATION

AN NDSA REPORT



December
2013

Results of a Survey of Organizations Preserving Digital
Content

Authors

Winston Atkins, Duke University Libraries

Andrea Goethals, Harvard Library

Carol Kussmann, Minnesota State Archives

Meg Phillips, National Archives and Records Administration

Mary Vardigan, Inter-university Consortium for Political and Social Research (ICPSR)

Representing the NDSA Standards and Practices Working Group

Contents

EXECUTIVE SUMMARY..... 2

ABOUT THE NATIONAL DIGITAL STEWARDSHIP ALLIANCE 2

INTRODUCTION 3

METHODOLOGY..... 3

 The Data Files.....4

 The Survey Codebook4

SURVEY CONTENT 5

FINDINGS 5

 Background Information6

 Digital Content7

 Digital Preservation Activities9

 Digital Preservation Organization and Staffing 13

 Staff Qualifications and Training 16

 Comments about Digital Preservation Program Staffing and Organization..... 18

ANALYSIS 19

RELATED WORK 21

NEXT STEPS..... 22

CONCLUSION 22

REFERENCES..... 23

APPENDIX A -- SURVEY QUESTIONS..... 24

APPENDIX B – POSTER: IPRES 2012 AT THE UNIVERSITY OF TORONTO 26



This work is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/).

Persistent URL: <http://hdl.loc.gov/loc.gdc/lcpub.2013655113.1>

EXECUTIVE SUMMARY

In 2012 the National Digital Stewardship Alliance (NDSA) Standards and Practices Working Group surveyed 85 institutions with a mandate to preserve digital content to investigate how such organizations staffed and organized their preservation functions. It was found that for most organizations surveyed, there was no dedicated digital preservation department to take the lead in this area so preservation tasks fell to a library, archive, or other department. Close to half of respondents thought that the digital preservation function in their organizations was well organized, but a third were not satisfied and many were unsure.

Addressing staffing levels, organizations expressed a preference to have almost twice the number of full-time equivalents (FTEs) that they currently had. While the survey showed that most organizations were retraining existing staff to manage digital preservation functions, questions were also asked about desired qualifications for new digital preservation managers. Having a passion and motivation for digital preservation and a knowledge of digital preservation standards, best practices, and tools were considered the most sought after skills.

Other findings from the survey showed that most organizations expected the size of their holdings to increase substantially in the next year with 20% expecting a doubling of content. Images and text files were the most common types of content being preserved. Most organizations were performing the majority of digital preservation activities in-house but many outsourced some activities (digitization was the most common) and were hoping to outsource more.

The survey provides some useful baseline data about staffing needs, and the NDSA Standards and Practices Working Group recommends that the survey be repeated in two to three years to show change over time as digital preservation programs mature and as more organizations self-identify as being engaged in digital preservation.

ABOUT THE NATIONAL DIGITAL STEWARDSHIP ALLIANCE

Founded in 2010, the National Digital Stewardship Alliance (NDSA) is a consortium of institutions that are committed to the long-term preservation of digital information. NDSA's mission is to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations. NDSA member institutions represent all sectors, and include universities, consortia, professional associations, commercial enterprises, and government agencies at the federal, state, and local levels.

More information about the NDSA is available from <http://www.digitalpreservation.gov/ndsa/>.

INTRODUCTION

Organizations establishing or scaling up digital preservation programs are faced with many staffing, scoping and organizational decisions. How many staff are needed and what kinds of skills, education and experience should they have? What types of positions should the institution create? Should it hire new staff or retrain existing staff? And how should the preservation program be scoped -- that is, what functions should be included directly in the program, provided by other parts of the organization, outsourced, or implemented through collaboration with other organizations? What organizational and staffing models work well?

Several members of the National Digital Stewardship Alliance (NDSA) Standards & Practices working group discovered that they shared a common interest in knowing how organizations were staffing and organizing digital preservation programs so that they could use the information to improve their own programs and were willing to work together on a survey. The group believed that the information would be valuable not only to its own members, but also to the broader NDSA and digital preservation communities.

The Preservation Staffing Survey, conducted by members of the NDSA Standards & Practices working group, was designed to shed light on how organizations responsible for digital preservation are addressing these staffing, scoping and organizational questions.

METHODOLOGY

To determine some of the approaches organizations used to staff and organize digital preservation programs, an ad hoc working group formed to develop and administer a survey. Over the course of two months in the spring of 2012, the working group members collaborated using Google docs to draft the survey questions. The survey was then transferred to Qualtrics¹, a web-based survey tool subscribed to by Harvard University, an NDSA member. Qualtrics provided the ability to publish previews of surveys, so working group members were able to review and provide additional feedback on the look and behavior of the survey as it would appear to a person filling out the survey.

The staffing survey was announced in late June 2012 through multiple national and international mailing lists associated with digital preservation and remained open through the beginning of September. It was open to any institution choosing to respond. The survey was started 131 times and completed 85 times for a 65% completion rate. Results were analyzed using the Qualtrics software during September 2012 and were presented as charts and graphs on a poster at iPRES 2012 in Toronto a month later (See Appendix B).

To more widely share the data files² in order for others to review or extend our findings, it was decided to make them available through the Inter-university Consortium for Political and Social Research (ICPSR), a digital repository that archives social and behavioral science data. The process involved exporting data out of Qualtrics and creating a codebook documenting the survey data. Since this project was conducted under

¹ Qualtrics. Qualtrics.com | Sophisticated Research Made Simple. <http://www.qualtrics.com>. Accessed September 26, 2013.

² National Digital Stewardship Alliance (NDSA). Standards and Practices Working Group. Survey of Staffing Practices and Needs Related to Digital Preservation, 2012. ICPSR34901-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-10-01. [doi:10.3886/ICPSR34901.v1](https://doi.org/10.3886/ICPSR34901.v1).

the auspices of the NDSA, the team felt it was important to demonstrate a commitment to digital preservation by properly archiving its own data. Details about the data files and resulting codebook are provided below.

The Data Files

The Qualtrics software provided several options for exporting the survey response data, including options for excluding responses that could reveal respondents' identities. Using this feature, we excluded responses for Questions 2, 21, and 22.³ After examining the exported data, we excluded responses for two additional questions because they contained information that could be used to identify the respondents' organizations: IP address (automatically supplied by Qualtrics), and Question 13.

We chose to deposit the data in two formats: CSV and SAV, which is a format compatible with SPSS.⁴ The CSV format provides ready access to the data and does not require specialized software. The SPSS-compatible format contained more robust metadata, which we felt could be useful.

An additional decision to make was which set of responses to deposit--those in which a respondent reached the final page and hit the 'Submit' button (85 responses), or the larger set of all surveys that were begun, whether or not the respondent submitted the responses (131 responses). We decided to include all the responses in the archived data, whether or not the respondent completed the survey, although only the completed responses were included in the survey analysis described in this paper.

The Inter-university Consortium for Political and Social Research's *Guide to Social Science Data Preparation and Archiving*⁵ helped us to assemble the necessary documentation for the survey. We also used ICPSR's guidelines on file formats to decide which formats we would use to deposit the data and the codebook.⁶

The Survey Codebook

The codebook establishes the context for the survey and its responses. It assumes that future users will have no prior knowledge of the survey or its data, so much of our effort went into ensuring that the Summary and Scope of Survey section conveyed the working group's mission and the survey goals, and provided an overview of our approach to inviting responses. The codebook also documents the full text of each question, the question's location in the table of responses, the possible responses, and the formats in which responses could be entered. In addition, we created an additional report using SPSS's "Codebook" report utility.

³ The complete survey is included in Appendix A.

⁴ SPSS statistical software: <http://www-01.ibm.com/software/analytics/spss/>. Accessed September 26, 2013.

⁵ Inter-university Consortium for Political and Social Research (ICPSR). *Guide to Social Science Data Preparation and Archiving: Best Practices Throughout the Data Life Cycle*. 5th Edition. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/>. Accessed August 28, 2013.

⁶ A codebook is "a document that provides information on the structure, contents, and layout of a data file." <http://www.ddialliance.org/bp/definition>. Accessed August 28, 2013.

If a respondent either quit the survey or chose not to answer a question, that decision was recorded with an empty cell or the value “-99”. An empty cell typically meant that the question was formatted for text input while the value “-99” indicated that the unanswered question was seen but not answered and formatted for a numeric response.

SURVEY CONTENT

The survey posed a series of questions related to the nature of the institution preserving content, the scope of the digital preservation program, and the number and types of positions responsible for carrying out digital preservation work. It also solicited information about the organizational structure of the program: whether there was a dedicated digital preservation department, and if not, whether the responsibility for this work was centralized in another department or decentralized in other parts of the institution. The survey also queried respondents about which functions were performed in-house and which were outsourced, and what the future plans for these functions were. Finally, to learn more about not just how organizations were approaching digital preservation but which approaches were perceived as working well, the survey probed for whether the institution was satisfied with the current organizational structure around preservation. There was also an option in the survey to reference or send related organizational charts or position descriptions.

FINDINGS

This section of the report includes the text of the original questions clustered into related themes, a presentation of the results for each question, and a short discussion of the implications of the results. A longer discussion of the larger picture of digital preservation staffing provided by the survey follows in the Analysis section.

The major themes are:

- Background Information
- Digital Content
- Digital Preservation Activities
- Digital Preservation Organization and Staffing
- Staff Qualifications and Training
- Comments about Digital Preservation Program Staffing and Organization

Since a major goal of the survey was finding out how repositories with different characteristics were currently staffed, whether they were satisfied with their current staffing, and what they would like their staffing model to look like in the future, questions found in several of the sections cover both “what we have” and “what we want.” For example, both the “Digital Preservation Organization and Staffing” and

“Staff Qualifications and Training” sections of this report shed light on both the present situation and the desired future in this way.

Background Information

After an introductory section (Q1), the survey started with general background questions to provide an understanding of the types of organizations that were responding to the survey.

Q2 (Required) - What is the name of your organization?

Q3 - Can we include the name of your organization in a list of organizations that responded to this survey? Knowing specific responding organizations may be helpful to people interpreting the survey results. If you agree to this we will still make our best effort to protect your individual survey responses so that no one will be able to connect your responses with you or your organization.

Q4 (Required) - Which of the following most closely describes the type or function of your organization?

Q5 - In which country do you reside?

Q22 - Please provide your contact information if you are willing to respond to follow-up questions.

The name of the organization (Q2) was required because it was important to know that there was only one response from each organization. These organizational names were not shared unless respondents specifically gave permission (Q3). Contact information was also requested for follow-up questions if organizations were willing to provide it (Q22).

As expected, the majority (86%) of the respondents were from the United States. Responses also came from Australia (1), Canada (4), Denmark (1), Estonia (1), Germany (1), Ireland (1), Japan (1), Netherlands (2), Switzerland (1), and United Kingdom of Great Britain and Northern Ireland (1).

As shown in Figure 1, the highest percentage of respondents represented academic libraries or archives (45%) followed by government entities (11%) and public libraries (10%). Fewer than 5% of the respondents were from independent libraries or archives; universities; academic institution departments other than a library or archive; research data repositories; non-profit organizations (not one of the listed types); national, federal, or legal deposit libraries; national libraries (mentioned under “Other”); national research infrastructure institutions (mentioned under “Other”); for-profit corporations, historical societies; botanic gardens (mentioned under “Other”); and public media institutions and trade associations (mentioned under “Other”).

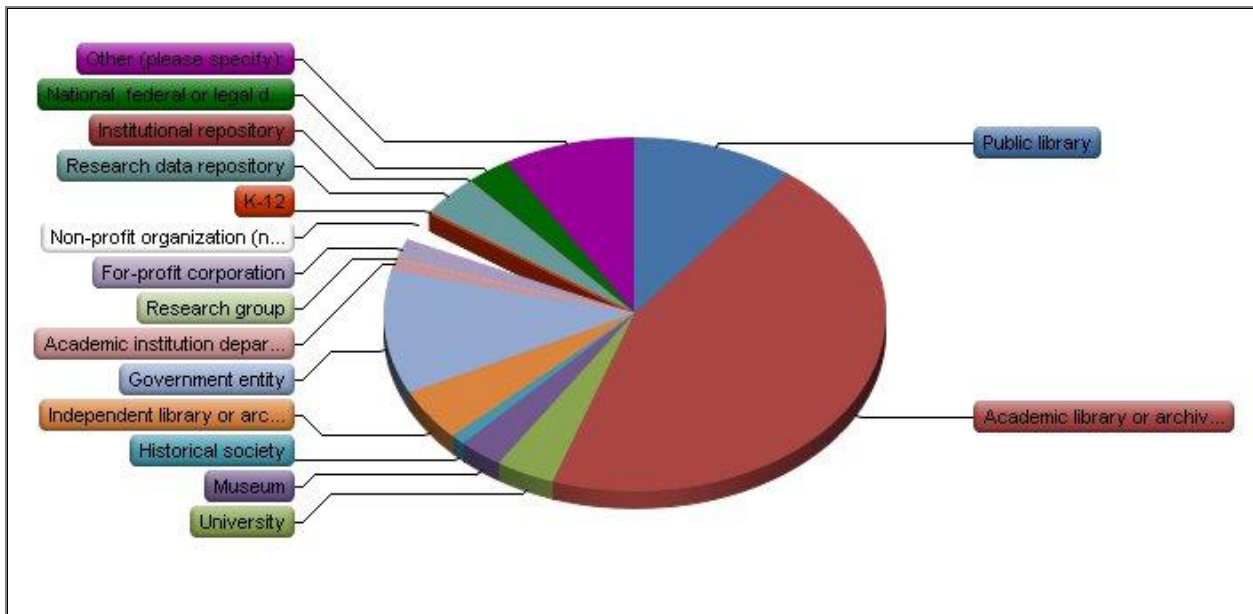


Figure 1: Types of organizations represented by respondents

Digital Content

Additional questions designed to elicit information about the scope of preservation efforts focused on the amount of digital content each organization was preserving and expected to be preserving in the near future.

Q6 - How much online or offline storage space are you using for your digital content, not including backup copies?

Q7 - What do you expect the percent of growth to be of your preserved digital content over the next year? *Please enter a number representing a percentage.*

Q8 - Roughly how much of each are you preserving in terms of number of files? (None, a little, some, a lot)

Survey results (see Figure 2) show that over half of the respondents (59%) were preserving 1-50 TB of digital content; 32% were preserving 51-500 TB; a small number (4%) were preserving over 500 TB; and a small number of respondents were not yet preserving digital content or were preserving over 500TB.

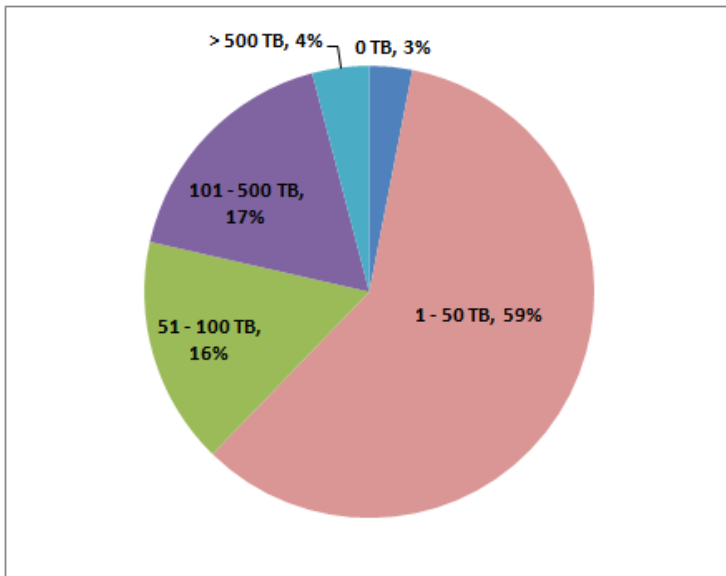


Figure 2: Amount of content organizations were preserving

While it is important in interpreting the survey results to know the amount of content organizations were preserving, it is also important to know the amount of content growth they were expecting in the near future. The survey (Q7) asked for an estimate of the percent growth of preserved digital content expected over the next year. As shown in Figure 3, 68% expected up to a 49% content increase, 12% expected a 50-99% content increase, and 20% expected at least a 100% content increase (at least a doubling of content). Of those expecting content to at least double, most organizations (76%) held 1-50TB of content.

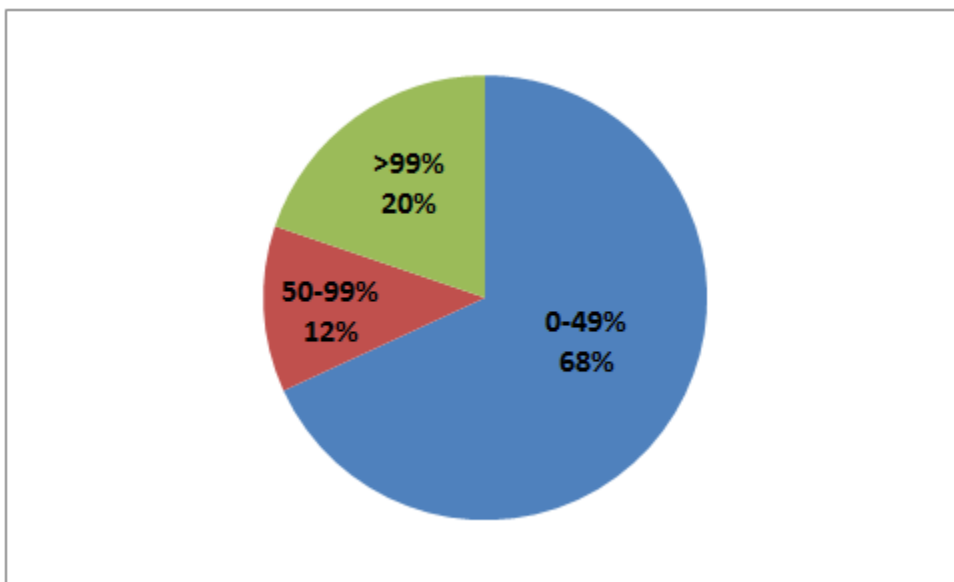


Figure 3: Expected percent growth in preserved content

It is also helpful to understand the types and variety of digital formats organizations were preserving. The survey asked respondents to specify and roughly quantify the formats they were preserving (Q8). The qualitative response options (“A lot, some, a little, none”) were chosen to elicit the overall dominance of types of content within each collection, regardless of the size of the collection or the exact number of files or percentage of storage space used for each. Figure 4 shows the results by response count. Images and text documents made up the majority of digital files being preserved; however, other formats such as audio and video, data sets, and GIS data collectively made up a large collection of digital content that must be managed by the responding organizations.

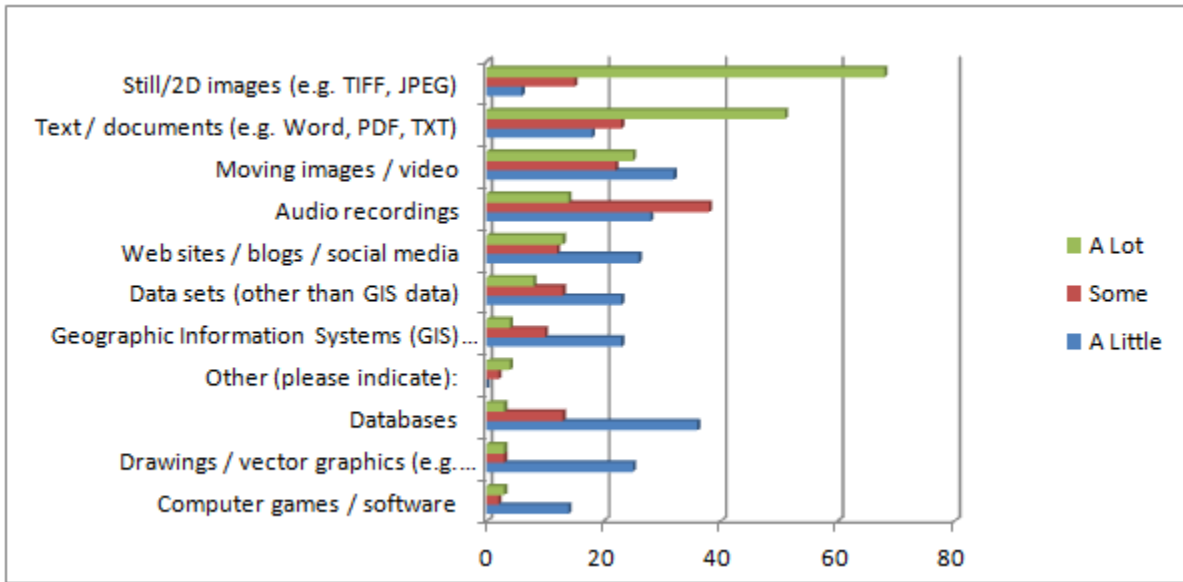


Figure 4: Types of content by response count (multiple items could be selected by each respondent)

Digital Preservation Activities

To understand how organizations defined digital preservation and how they managed these functions, the survey asked questions about participation in cooperative efforts, the types of activities respondents considered in scope for digital preservation, and which of these were done in-house or outsourced.

Q9 - Do you participate in any digital preservation consortial or cooperative efforts?

Q10 - Which of these activities are considered part of the scope of the digital preservation function at your organization, whether or not you have implemented this activity yet? Check all that apply.

Q11 - Which of these activities are you currently doing in-house and/or outsourcing? If you haven't implemented an activity yet, leave it unchecked.

Q12 - Which of these activities would you like to outsource? Also include any activities that you are already outsourcing and would like to continue outsourcing.

The survey asked respondents about what activities each organization considered part of digital preservation within their organization (Q10). Choices included:

- Selection for preservation
- Digitization
- Metadata creation/extraction
- Descriptive cataloging
- Transformation / migration of formats
- Creation of access copies
- Normalization of files
- Fixity checks
- File format identification
- File format validation
- Emulation
- Content replication
- Secure storage management
- Technology watch
- Development and maintenance of tools
- Preservation planning
- Development of preservation policies and strategy
- Development of guidelines for content creators
- Research
- Preservation education, training and outreach

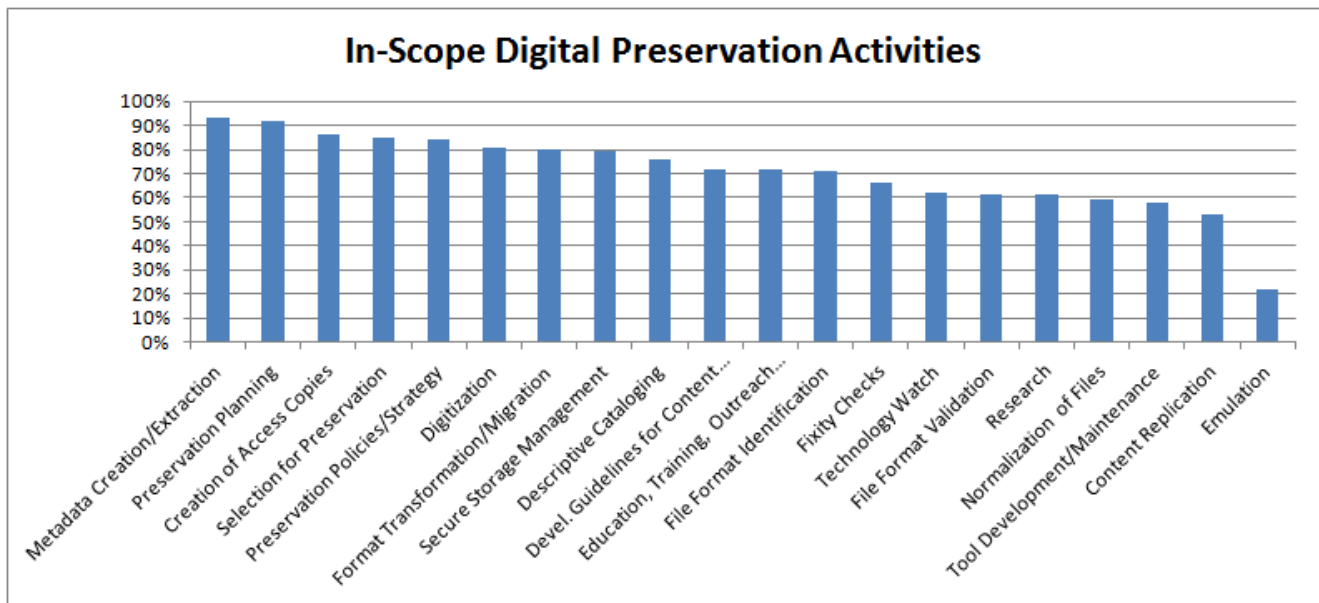


Figure 5: In-scope digital preservation activities

Figure 5 shows that everything on the list was in scope for over half of the respondents with the exception of emulation, which was only in scope for 22%. This demonstrates the complexity of digital preservation and its multi-faceted nature, underscoring the fact that a variety of skills are needed to address digital preservation issues.

While Q10 asked which activities respondents considered to be part of digital preservation, a follow-up question (Q11) asked respondents which of these activities were currently being done in-house or outsourced (note that respondents were only presented with the options that they selected in Q10 as part of the digital preservation scope at their institutions).

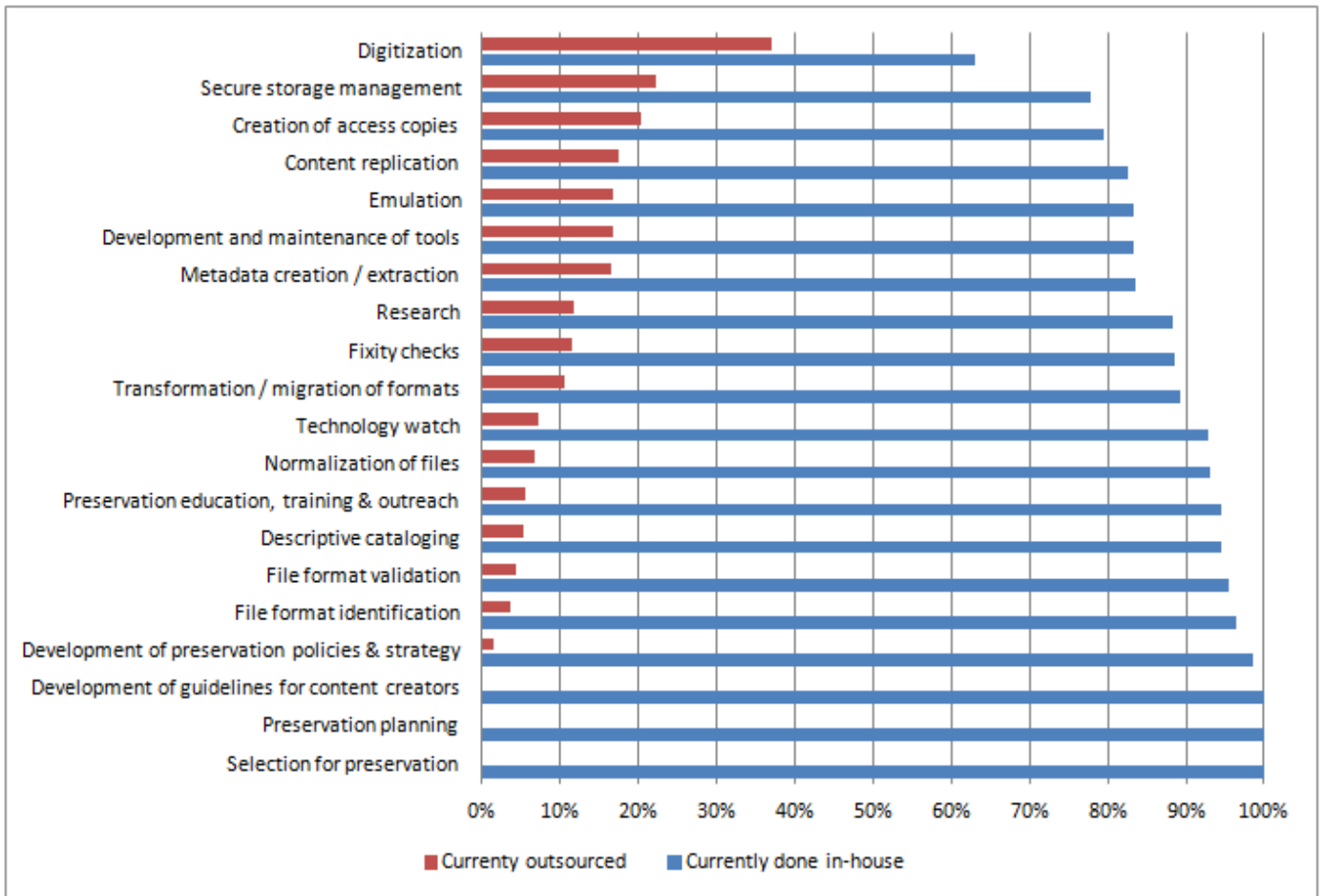


Figure 6: Percent of digital preservation activities done in-house and outsourced

As Figure 6 shows, the majority of digital preservation functions were being done in-house. Where these functions were in scope for digital preservation at respondents’ institutions, 100% were conducting preservation planning, developing guidelines for content creators, and making selections for preservation in-house. Over 90% were also conducting these functions in-house: developing preservation policies and

strategies; file format identification, validation and normalization; descriptive cataloging; preservation education, training and outreach; and technology watches.

Not surprisingly, the highest function being outsourced was digitization (37%), although 63% were also doing it in-house. Other functions outsourced by more than 15% of the respondents for this question included secure storage management (22%), creation of access copies (20%), content replication (17%), development and maintenance of tools (17%), emulation (17%) and metadata creation/extraction (16%).

When respondents were asked which functions their organizations would *like* to outsource (Q12), including functions already being outsourced that would continue to be outsourced, digitization (65%), secure storage management (43%), file format migration/transformations (34%), metadata extraction (32%), and the development and maintenance of tools (29%) were at the top of the list (see Figure 7). As before, respondents were only presented with the options that they selected in Q10 as in scope for the digital preservation function at their institutions.

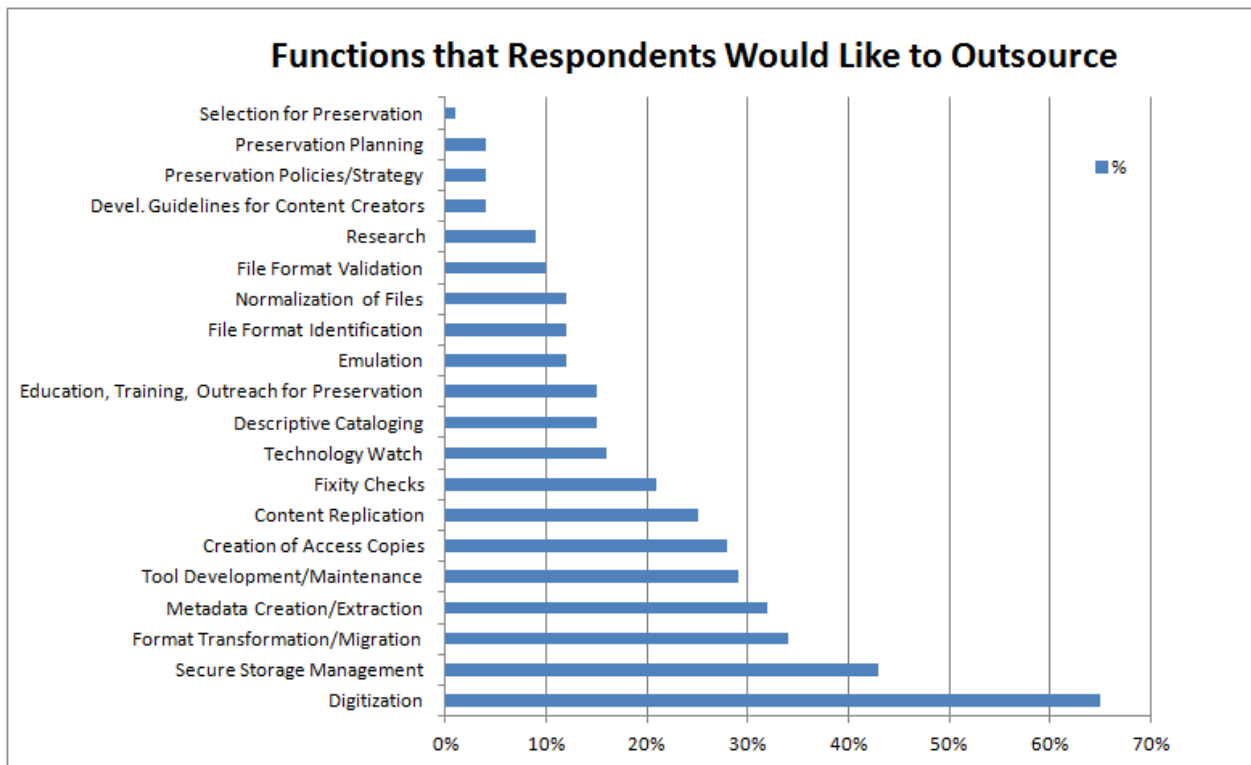


Figure 7: Functions respondents would like to outsource

Respondents were also asked if they participated in digital preservation consortia or other cooperative efforts (Q9). This response was a 54%/46% yes/no split. Those who answered yes were asked to list the name of the consortium or cooperative. Answers included national, regional and local organizations such as HathiTrust, LOCKSS, MetaArchive, NDSA/NDIIPP, PeDALS, local digital libraries and many more organizations and initiatives. Some were content-specific while others were more general.

Participation in consortia did not seem to be correlated with the amount of content an organization was preserving. Results show that of organizations storing between 101-500TB, eleven participated in consortia and six did not, while organizations with under 101TB or over 500TB had roughly the same rate of participation as non-participation in consortia. It would be interesting to know if organizations believe that their participation in consortia or cooperative efforts makes them better prepared to care for digital materials.

Digital Preservation Organization and Staffing

Questions were asked to determine if organizations had dedicated digital preservation departments within their organizations, which department(s) took the lead with digital preservation, and if they were satisfied with the way digital preservation functions were organized within their organizations.

Q13 - Is there a dedicated digital preservation department within your organization? If yes, please indicate the name of the department.

Q14 - Which department(s) take the lead for digital preservation within your organization? If this is a fairly equally distributed effort choose more than one.

Q16 - The way our digital preservation function is currently organized (staffing levels, expertise, where they are placed within the larger organization) works well.

Responding to Q13, only 33% of respondents had dedicated digital preservation departments within their organizations, while 59% did not. Eight percent responded that the question was not applicable to their situations.

When asked which department(s) took the lead for digital preservation within their organizations (Q14), 73% responded that these duties fell most often to a library, archive or other department that stewarded the collections, and 42% responded that it was an IT department, as shown in Figure 8. Note that respondents could select more than one option for this question.

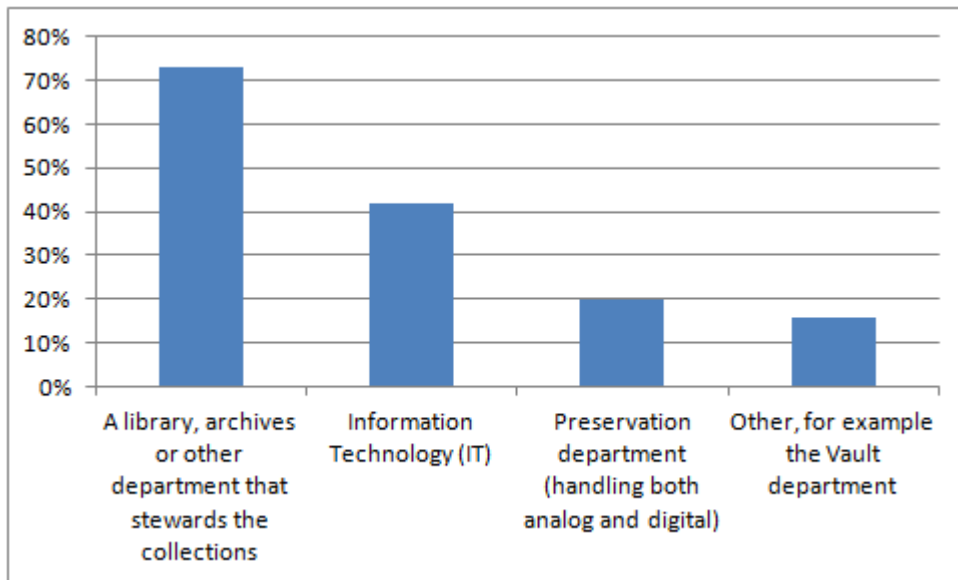


Figure 8: Department(s) leading preservation efforts

Q16 asked if the way the institution’s digital preservation function was currently organized (staffing levels, expertise, where they were placed within the larger organization) worked well. The good news is that 43% agreed or strongly agreed that the digital preservation function within their organizations was well organized; however, a significant number, 34%, were not satisfied with how things were organized, and 23% were unsure (answering “Neither Agree nor Disagree”), as shown in Figure 9.

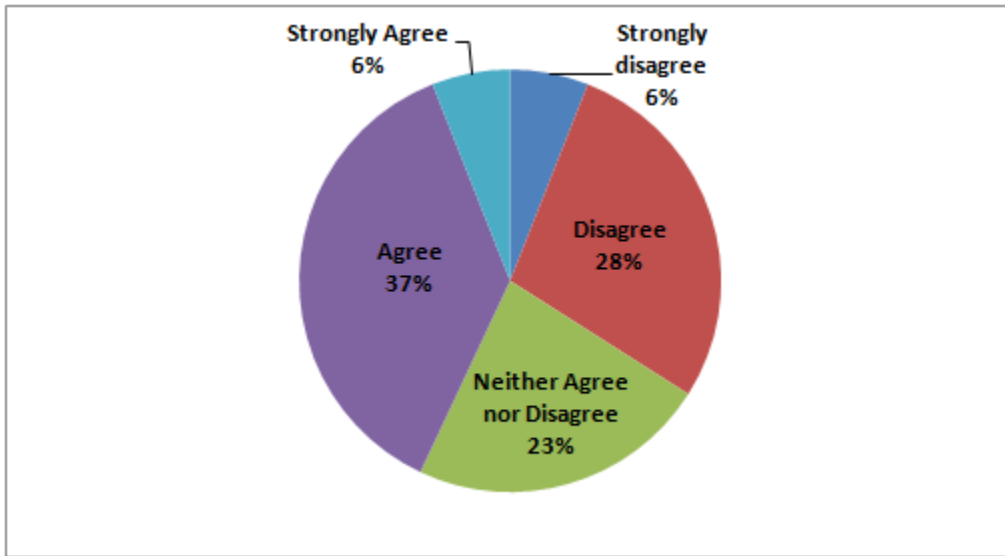


Figure 9: Level of agreement that the digital preservation function was organized well within the organization

To better understand how organizations staffed themselves in regards to digital preservation, organizations were asked how many full-time equivalents (FTEs) they had for specific positions or roles relating to digital preservation as well as what their ideal number of FTEs for each role would be. The difference between these two values shows how current staffing levels compared to respondents’ needs/wants.

Q15 - For each of these positions, how many FTE do you currently have doing digital preservation in your organization, and how many would be ideal? FTE stands for Full-time equivalent. For example a 1.0 FTE could mean one person full-time or 2 people half-time; a 0.5 FTE could mean one person half time or 2 people quarter-time. *Do not include people doing work other than digital preservation. Please use whole numbers or decimals as appropriate.*

In evaluating the responses, it was found that organizations would like to have almost twice as many FTEs as they currently had working on digital preservation activities. Specific results for each position are shown in Figure 10. (Note: The graph below only includes responses for which there were no blank answers. For example, if a response for current FTEs was ‘2’ and the field for ideal FTEs was left blank, we could not make assumptions regarding whether the respondent was happy with the two FTEs and the ideal value should be the same or whether the respondent wanted more or less people. Similarly if a respondent filled

in the ideal FTEs but left the current FTEs column blank, it was not recorded in the graph below; we could not assume a blank answer was zero.)

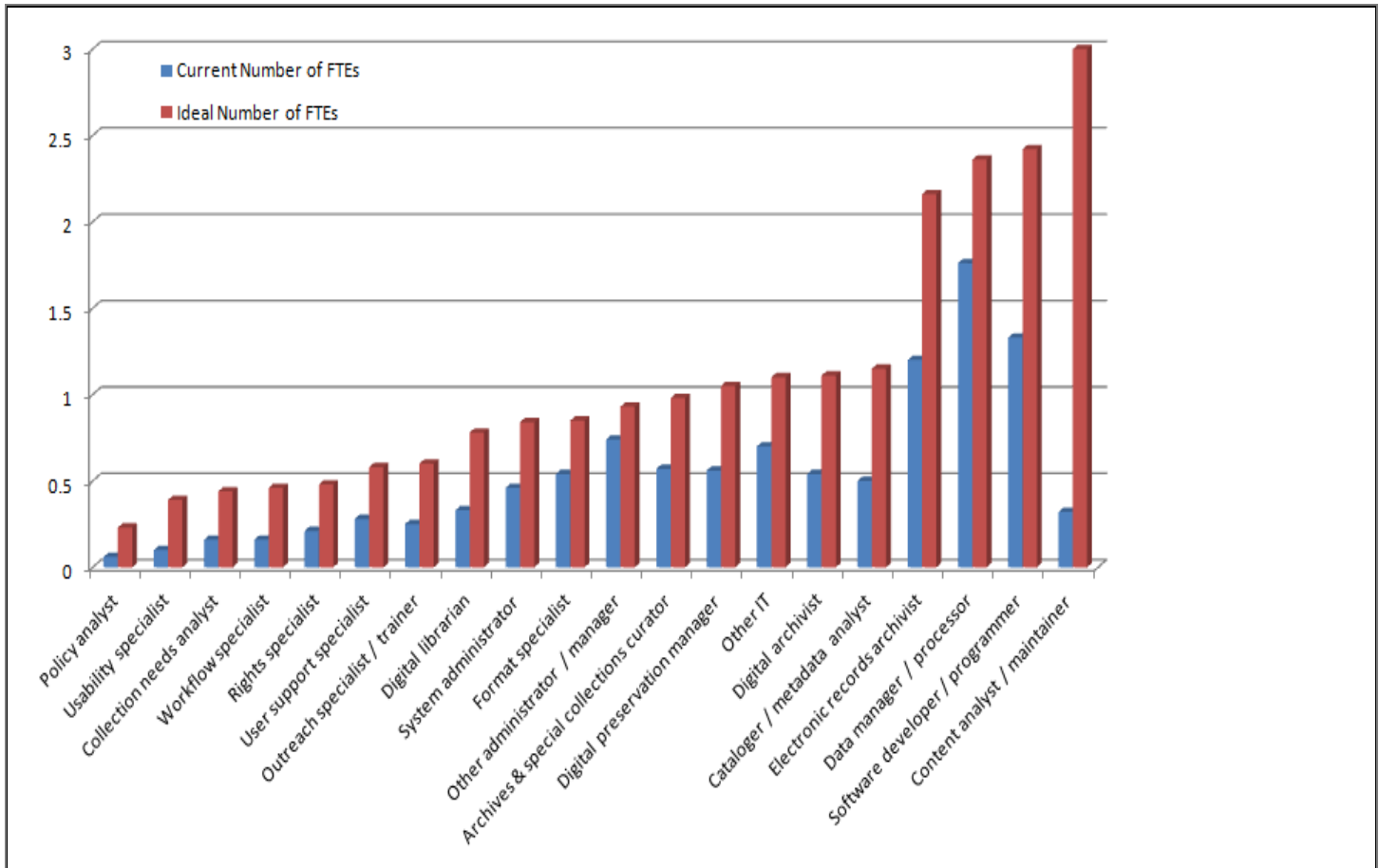


Figure 10: Current vs. ideal staffing

To learn about how and where the digital preservation function fit within organizations’ unique structures, respondents were asked to share organizational charts and/or applicable position descriptions:

Q20 - Do you have organizational charts or position descriptions that you’d be willing to share? Any documents you share would provide context to your answers and would be kept private to the NDSA Standards and Practices Working Group unless you give us explicit permission in the future to share more broadly.

Q21 - If they are on-line, what are the URL(S)? If they are not on-line, please email this supplementary data to **nds@loc.gov** with the subject line **Staffing Survey**

In response to Q20, sixteen organizations indicated that they had and were willing to share this type of information. In response to Q21, eight either provided links or responded that they would send the

information. The small number of responses may show that the documentation of digital preservation processes has yet to be formalized and that institutions are still developing these as two-thirds of survey respondents were willing to share their own contact information for follow-up questions.

Staff Qualifications and Training

Questions 17 and 18 asked for the desired qualifications of digital preservation staff, and whether organizations hired new staff with digital preservation experience or retrained existing employees to manage digital preservation functions.

Q17- For in-house staff, did you hire experienced digital preservation specialists and/or retrain existing staff? Check all that apply.

Q18 - Please rate the importance of each of these items if you were hiring a new digital preservation manager at your organization.

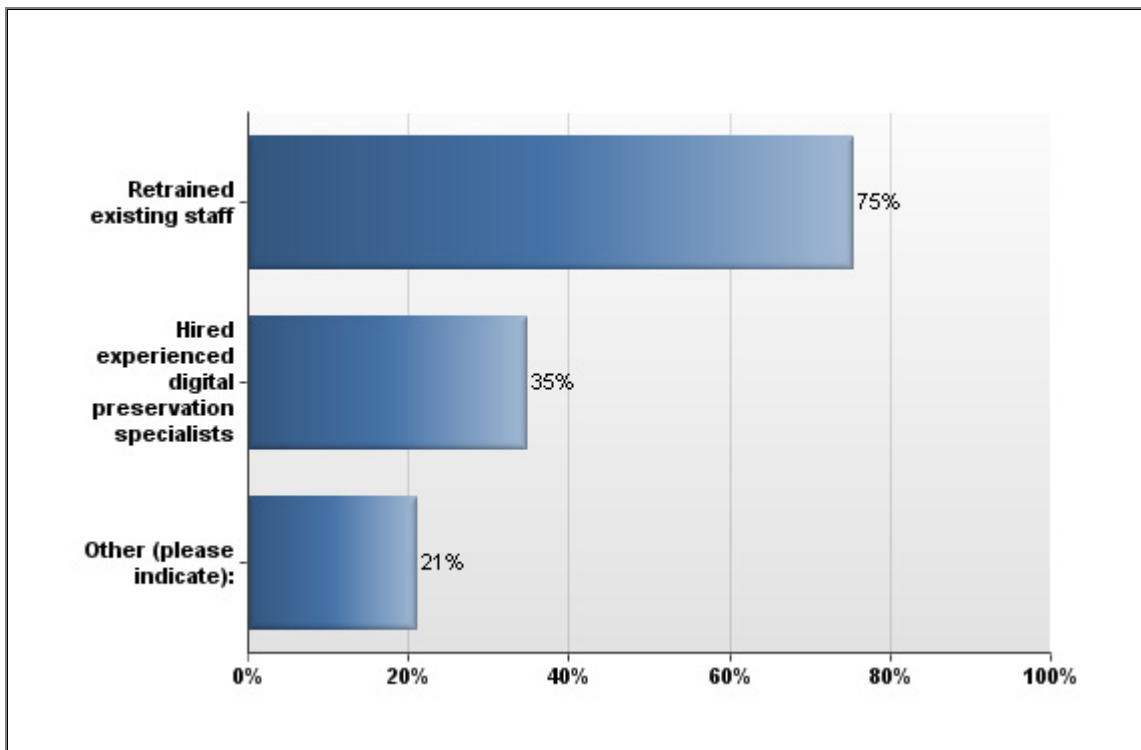


Figure 11: Hiring new staff vs. training existing staff

Understanding that respondents could select more than one answer, we found that 75% retrained existing staff, 35% hired experienced digital preservation specialists, and 21% selected “other” (see Figure 11). Explanations in the “other” category included hiring new staff with competence or other experience and

training them on digital preservation skills, hiring specialists to train staff, and just learning the required skills as the processes moved forward.

Given the chance to hire a new digital preservation manager, organizations were asked to rank the relative importance of skills, knowledge and education. They could be ranked as extremely important, very important, somewhat important, somewhat unimportant, very unimportant, or not at all important.

A passion and motivation for digital preservation and knowledge of digital preservation standards, best practices and tools were considered the most sought after skills, followed by general communication and analytical skills. Respondents were less concerned with the specific degrees or certificates people held. Figures 12 and 13 below break out the results by what respondents considered the most important and the least important qualifications for a digital preservation manager.

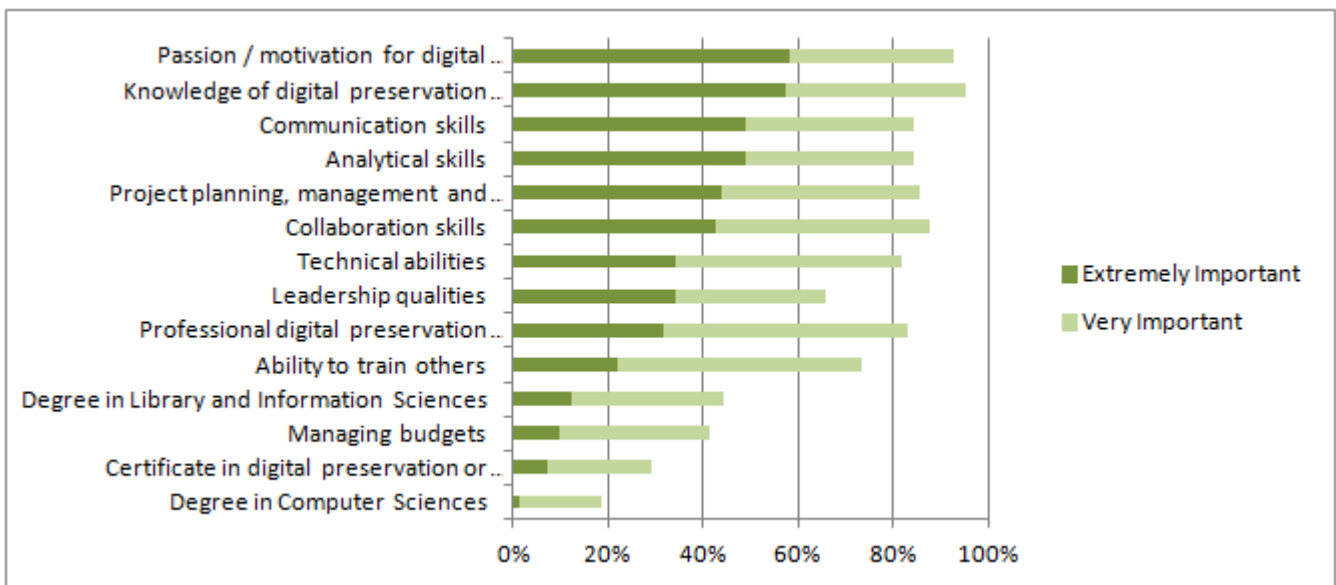


Figure 12: Most important qualifications by percentage of respondents

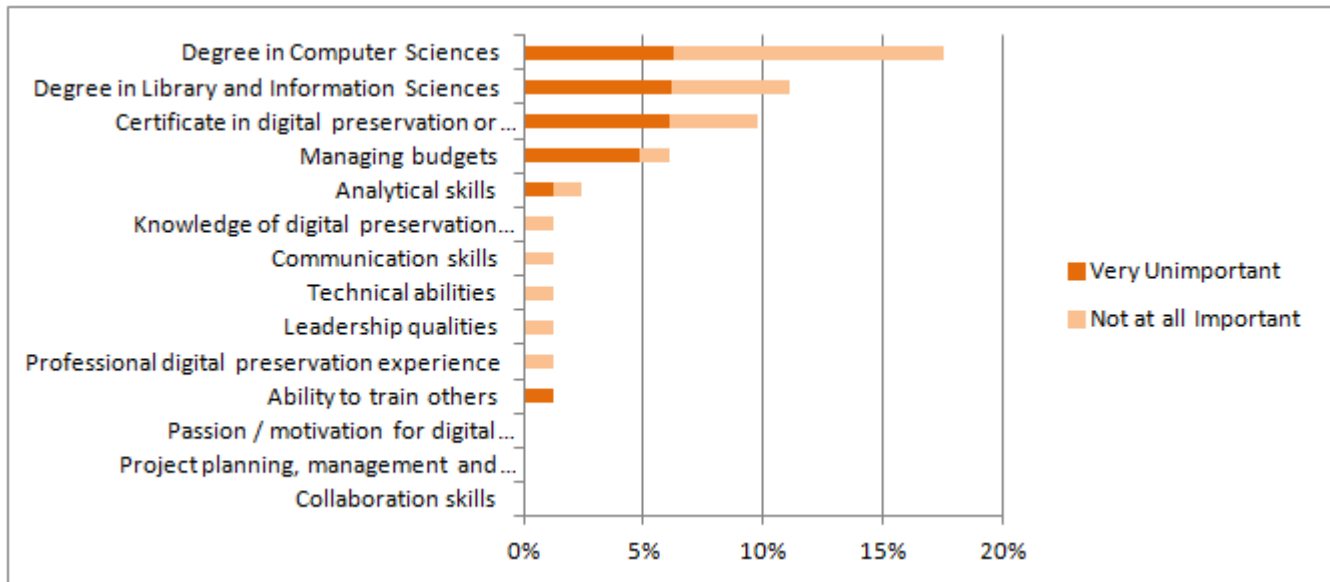


Figure 13: Least important qualifications by percentage of respondents

Comments about Digital Preservation Program Staffing and Organization

To give respondents a chance to voice their opinion, clarify results, or explain their situations, an open-ended question was asked about digital preservation programs. Twenty-five organizations responded with additional comments.

Q19 - Is there anything else you'd like to share about the way you think an effective digital preservation program should be staffed and organized?

Collaboration, both within an organization and externally, was the most frequently discussed topic. Many organizations were looking for examples from others who had already addressed the issues they were facing. Sharing experiences can be a first step in collaborating.

“Smaller organizations will depend on internal and external collaborations.”

“Collaborative, multi-institutional efforts will be increasingly cost effective as content demands scale up.”

Not surprisingly, responses also emphasized the importance of buy-in from the entire organization and budget sustainability.

“There is a general lack of understanding of the budgetary demands on digital collections and preservation that equal, and usually exceed, traditional collection development. The “lights-on” cost are rarely, if ever, discussed or budgeted, thus digital libraries and preservations typically are funded with leftovers.”

Comments specifically about staffing focused on hiring or having a main person in charge of digital preservation with whom people performing digital preservation tasks would all work. In many cases these other people would be in various departments across the organization.

“I don’t think this needs to be a big staff working solely on digital preservation, but have one person full time and then have systems and digital collections people coordinating with that person.”

Other responses focused on the challenges of deciding how to organize staff as digital preservation programs at organizations develop. One respondent said “there is a significantly different staffing level for project start up and ongoing maintenance.” Another said “Our digital preservation efforts are still in their infancy. It is therefore quite difficult to gauge what our future staffing needs will be. We recently hired a Digital Preservation Librarian and our responses would likely be quite different one year from now. Future staffing needs will depend on the systems and tools we decide to implement, and the buy-in we are able to obtain from administrators in support of our long-term goals.”

The resulting themes and similarity of the comments show that organizations are not alone in their concerns and challenges.

ANALYSIS

Overall, the results of the Digital Preservation Staffing survey indicate that organizations are making do with what they have and generally think that their digital preservation programs and staffing are working well, but they feel a distinct need for more people to help do the work. It would be interesting to compare the number of additional people desired for the digital preservation function with the additional people desired for other functions in these organizations. Is digital preservation more understaffed than other functions? Is the pool of qualified applicants smaller? It was also interesting that relatively few organizations had a dedicated digital preservation unit. In most cases, digital preservation was a shared responsibility or was assigned to a custodial unit like a library or archives in a larger institution.

There was an interesting relationship between the responses to different questions. For example, the organizations that reported the greatest anticipated increase in the size of their holdings ($\geq 50\%$) over the next year (Q7) were the least happy with their current staffing (Q16) as shown in Figure 14. One imagines that these organizations anticipate a significant increase in workload and are concerned about how they will do the work with existing staff.

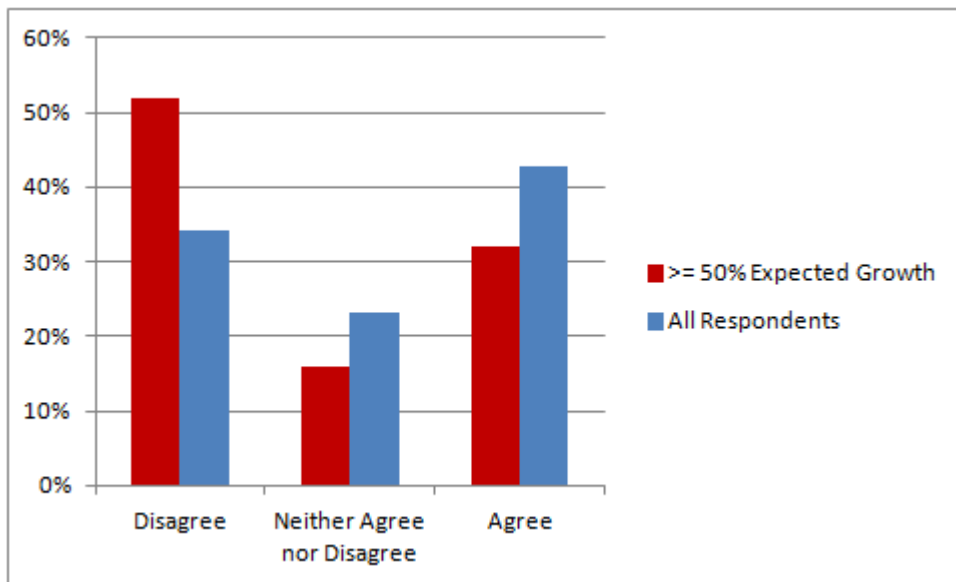


Figure 14: Difference in satisfaction with staffing between the institutions expecting a very high content growth rate and all survey respondents [In response to Q16, which asked organizations to rank “The way our digital preservation function is currently organized (staffing levels, expertise, where they are placed within the larger organization) works well.”]

The fairly high percentage of respondents who answered that their current digital preservation organization was neither working well nor not working well (i.e. answered ‘neither agree nor disagree’ to Q16) begs for further work. It is likely that this response reflects several different situations: some organizations may have chosen the options because some things worked well and others didn’t, some because they weren’t sure yet what success looks like, and some because their current situation was passable, but neither really good nor really bad.

It is difficult to know exactly how to interpret the interrelated responses to Q10, Q11, and Q12. Question 10 asked whether a long list of functions was considered in scope for digital preservation, and Q11 and Q12 asked which functions were done in house or outsourced, or which the organization would like to outsource. The phrasing of Q10 put the focus on how the organization defined the digital preservation function, not which of those functions they considered important or which they performed. For example, in some organizations, the creation of access copies might be considered an access function rather than a preservation function. It might have been interesting for the survey to elicit whether that function was done in-house or outsourced, regardless of whether or not it was considered part of digital preservation.

The responses about what should be done in-house versus outsourced help indicate what are considered core preservation functions that must be the responsibility of the organization itself. In general, planning and policymaking activities are least likely to be outsourced; they are viewed as core functions of the organization, which makes sense. Implementation tasks involved in carrying out preservation policies and plans, while still done often in-house, are more likely to be outsourced as these processes can often be standardized and perhaps done more cheaply by specialists. For example, the task of digitization is

currently done in-house by over 90% of respondents, but outsourcing digitization is also well accepted; outsourcing storage management is desired by over 40% of organizations.

The skill set desired for a digital preservation manager has interesting implications for digital preservation education. The person in charge of digital preservation at an institution should be committed and knowledgeable about standards, but many of their highly desired skills are the skills of any effective manager and collaborator. The survey did not ask about the desired skills for lower level positions; presumably the mix of managerial and technical skills would be different for more technical positions.

RELATED WORK

Over the past few years several surveys and studies have sought to answer questions similar to those asked in the NDSA Staffing Survey.

In Europe, a survey by Engelhardt et al.⁷ for the Digital Curation Vocational Education Europe (DigCurV) project focused on digital preservation and curation training needs. It included questions about the number of digital preservation staff in the organization, hiring of new digital preservation staff, training existing staff, and the relative importance of particular skills. The majority of the respondents came from Europe: only 56 respondents came from the US. In comparison, the majority of the NDSA staffing survey's respondents were from the US.

Similarly, a recent survey conducted by the Alliance for Permanent Access to the Records of Science Network (APARSEN)⁸ included a question on the number of people doing digital preservation work. Their survey was aimed at research libraries in Europe, while the NDSA staffing survey was targeted toward many different types of organizations, and primarily in the US.

A survey by Nelson et al.⁹ for the Association of Research Libraries (ARL) asked about the number of staff and how new staff were identified or recruited. However, the focus of their query was different. The ARL questions applied to staff collecting as well as managing digital material, and that survey restricted the material to born-digital content. The NDSA staffing survey focused on digital preservation staffing only and applied to all digital content the organization was preserving, not just born-digital content.

Finally Kim et al.¹⁰ investigated job announcements within academic libraries to identify competencies required for digital curation positions.

⁷ Engelhardt, C., Strathmann, S. & McCadden, K. (2012). Report and analysis of the survey of training needs. Retrieved 6/8, 2012, from <http://www.digcur-education.org/eng/Resources/Report-and-analysis-on-the-training-needs-survey>.

⁸ APARSEN (2013). D36.1 Business Preparedness Report. Retrieved April 23, 2012, from http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/03/APARSEN-REP-D36_1-01-1_0.pdf

⁹ Nelson, N., Shaw, S., Ghering, C., Schmidt, L., Belden, M., Esposito, J., Pyatt, T., Deromedi, N. and Shallcross, M. (2012) SPEC Survey on Managing Born Digital Special Collections and Archival Materials. Retrieved June 8, 2012, from <http://www.arl.org/stats/specsurveys/index.shtml>

¹⁰ Kim, J., Warga, E., and Moen W. (2012). Digital curation in the academic library job market. ASIST 2012, October 28-31, 2012; Baltimore, MD. Retrieved April 23, 2012, from <https://www.asis.org/asist2012/proceedings/Submissions/283.pdf>

While these surveys differ from the NDSA survey in various ways, it would be interesting to do a more in-depth analysis of the findings for comparison purposes.

NEXT STEPS

The NDSA Digital Preservation Staffing Survey project team is planning several activities to wrap up its work on this survey and follow up on the information gathered.

The project team is encouraging reuse of the survey questions in other environments. Recognizing the preponderance of United States respondents to the NDSA survey, it would be worthwhile to circulate it among other preservation communities, for example in Europe, so that the responses could be compared.

The project team plans to repeat the survey in a couple of years to see if any change can be detected as digital preservation programs mature and as more organizations self-identify as being engaged in digital preservation. (The increasing membership in the NDSA itself over the last year is an indication that the next survey may reach more organizations.)

While repeating questions exactly will guarantee that responses to a question over time are comparable, there are a few questions that the team will consider clarifying even at the risk of muddying future comparison. For example, the choices available in Q4 to describe the type of organizations responding to the survey were apparently not clear or mutually exclusive. A number of organizations choose “other” when the team believed they could have chosen a listed category. It may also be helpful to clarify the intent of Q10 about the scope of digital preservation in an organization, and the relationship between Q10 and Q11 and Q12 with respect to outsourcing.

Perhaps the team will also follow up on functions that an organization performs, but that are not considered within the domain of digital preservation in that organization. The survey collected several comments from respondents indicating that it was difficult for them to answer the question about the roles and number of full-time equivalent staff doing digital preservation. Some said, for example, that because titles are not uniformly related to tasks across institutions, the same basic job could be done by people with quite different titles. Similarly, if a job was performed but not considered in scope of the digital preservation function, how should an organization answer this question?

CONCLUSION

The NDSA Digital Preservation Staffing Survey established a useful baseline description of how 85 organizations involved in digital preservation in 2012, mainly in the United States, addressed staffing, scoping and organizational questions. In particular, the survey answered questions about how institutions defined the scope of the digital preservation function, organized and hired or trained staff, and conducted their work in-house or through outsourcing or collaborative relationships. Although this survey only asked about types and volume of digital material being preserved to set the context for staffing and organizational questions, the nature of the collections and their expected growth rate are significant

findings in themselves. Future surveys can follow up on the 2012 survey to look for trends over time as the field and digital preservation programs mature.

REFERENCES

- APARSEN (2013). D36.1 Business Preparedness Report. Retrieved April 23, 2012, from http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/03/APARSEN-REP-D36_1-01-1_0.pdf
- Engelhardt, C., Strathmann, S. & McCadden, K. (2012). Report and analysis of the survey of training needs. Retrieved 6/8, 2012, from <http://www.digcur-education.org/eng/Resources/Report-and-analysis-on-the-training-needs-survey>.
- Kim, J., Warga, E., and Moen W. (2012). Digital curation in the academic library job market. ASIST 2012, October 28-31, 2012; Baltimore, MD. Retrieved April 23, 2012, from <https://www.asis.org/asist2012/proceedings/Submissions/283.pdf>
- NDSA Standards and Practices Working Group (2012). "Who's Minding the (Data) Store? Results of the NDSA Digital Preservation Staffing Survey." Poster presented at iPRES 2012 at the University of Toronto <http://www.digitalpreservation.gov/ndsa/documents/NDSA-staff-survey-poster-ipres2012.pdf>
- National Digital Stewardship Alliance (NSDA). Standards and Practices Working Group. Survey of Staffing Practices and Needs Related to Digital Preservation, 2012. ICPSR34901-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-10-01. [doi:10.3886/ICPSR34901.v1](https://doi.org/10.3886/ICPSR34901.v1)
- Nelson, N., Shaw, S., Ghering, C., Schmidt, L., Belden, M., Esposito, J., Pyatt, T., Deromedi, N. and Shallcross, M. (2012). SPEC Survey on Managing Born Digital Special Collections and Archival Materials. Retrieved June 8, 2012, from <http://www.arl.org/stats/specsurveys/index.shtml>
- Pilette, R. (2012, June 6). digital preservation position [Electronic mailing list message] digital-preservation@jiscmail.ac.uk. Retrieved from <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=digital-preservation;c0dea43.1206>
- Qualtrics. Qualtrics.com | Sophisticated Research Made Simple. <http://www.qualtrics.com>. Accessed September 26, 2013.
- SPSS statistical software: <http://www-01.ibm.com/software/analytics/spss/>. Accessed September 26, 2013.

APPENDIX A -- SURVEY QUESTIONS

Q1. This survey was created by the Standards and Practices Working Group of the National Digital Stewardship Alliance (NDSA). If you would like to learn about the NDSA, including how your institution can become a member, please see the NDSA web site (<http://www.digitalpreservation.gov/ndsa/>).

This survey is intended for organizations that are currently responsible for digital preservation, whether that responsibility is fulfilled in-house or outsourced to a commercial, non-profit, or consortial provider. It will be used to understand current and ideal practice in staffing digital preservation programs.

Only one response should be submitted per organization.

We will make our best effort to protect your individual survey responses so that no one will be able to connect your responses with you or your organization. Any personal information that could identify you or your organization will be removed or changed before results are made public. We will combine your responses with the responses of others and make the aggregated results public, and preserve the anonymous data long-term for research purposes.

Please email any questions to ndsa@loc.gov with the subject line Staffing Survey.

All questions are optional unless otherwise noted.

Q2. (Required) What is the name of your organization?

Q3. Can we include the name of your organization in a list of organizations that responded to this survey? Knowing specific responding organizations may be helpful to people interpreting the survey results.

If you agree to this we will still make our best effort to protect your individual survey responses so that no one will be able to connect your responses with you or your organization.

Q4. (Required) Which of the following most closely describes the type or function of your organization?

Q5. In which country do you reside?

Q6. How much online or offline storage space are you using for your digital content, not including backup copies?

Q7. What do you expect the percent of growth to be of your preserved digital content over the next year? Please enter a number representing a percentage.

Q8. Roughly how much of each are you preserving in terms of number of files?

Q9. Do you participate in any digital preservation consortial or cooperative efforts?

Q10. Which of these activities are considered part of the scope of the digital preservation function at your organization, whether or not you have implemented this activity yet? Check all that apply.

Q11. Which of these activities are you currently doing in-house and/or outsourcing? If you haven't implemented an activity yet, leave it unchecked.

Q12. Which of these activities would you like to outsource? Also include any activities that you are already outsourcing and would like to continue outsourcing.

Q13. Is there a dedicated digital preservation department within your organization?

Q14. Which department(s) take the lead for digital preservation within your organization? If this is a fairly equally distributed effort choose more than one.

Q15. For each of these positions, how many FTE do you currently have doing digital preservation in your organization, and how many would be ideal? FTE stands for Full-time equivalent. For example a 1.0 FTE could mean one person full-time or 2 people half-time; a 0.5 FTE could mean one person half time or 2 people quarter-time. Do not include people doing work other than digital preservation. Please use whole numbers or decimals as appropriate.

Q16. The way our digital preservation function is currently organized (staffing levels, expertise, where they are placed within the larger organization) works well.

Q17. For in-house staff, did you hire experienced digital preservation specialists and/or retrain existing staff? Check all that apply.

Q18. Please rate the importance of each of these items if you were hiring a new digital preservation manager at your organization.

Q19. Is there anything else you'd like to share about the way you think an effective digital preservation program should be staffed and organized?

Q20. Do you have organizational charts or position descriptions that you'd be willing to share? Any documents you share would provide context to your answers and would be kept private to the NDSA Standards and Practices Working Group unless you give us explicit permission in the future to share more broadly.

Q21. If they are on-line, what are the URL(s)? If they are not on-line, please email this supplementary data to nds@loc.gov with the subject line Staffing Survey

Q22. Please provide your contact information if you are willing to respond to follow-up questions.

APPENDIX B – POSTER: IPRES 2012 AT THE UNIVERSITY OF TORONTO

Who's Minding the (Data) Store?

Results of the NDSA Digital Preservation Staffing Survey

<http://www.digitalpreservation.gov/nlsa/index.html>

Andrea Goethals, Harvard University, andrea_goethals@harvard.edu
Jimi Jones, Library of Congress, jjones@loc.gov
Carol Kusmann, Minnesota Historical Society / State Archives, carol.kusmann@mnhs.org
Kate Murray, National Archives and Records Administration, kate.murray1@nara.gov
Meg Phillips, National Archives and Records Administration, meg.phillips@nara.gov

Survey Goals

Businesses, cultural memory institutions and government bodies that want to responsibly preserve digital assets face significant staffing challenges. The National Digital Stewardship Alliance (NDSA) Standards & Practices Working Group conducted a survey of organizations responsible for digital preservation to gain insight into how organizations worldwide are addressing staffing, scoping and organizational questions.

Respondent Characteristics

Completion Rate
The survey was started 131 times, and completed 85 times (a 65% completion rate).

Respondent Location
86% of the respondents were based in the United States, but 0% came from Europe, 4% from Canada, and 1% each from Japan and Australia.

Institution Type

47% (highest percentage) of the respondents were academic libraries or archives followed by government entities (11%) and public libraries (10%).

Content Type

Still images are the most common content type being preserved, followed by textual documents. As reported, databases and computer games/software are among the least preserved content type.

Data Storage Amount

Most respondents (60%) are preserving 50 TB or less while 36% are storing over 50 TB. A small percentage (4%) are storing more than 500 TB.

Expected Data Growth

21% of respondents expect a 50% or more increase in the amount of digital content in the next year.

What We Have

Where is In-House Work Done?
Libraries & Archives tend to take the digital preservation lead within our institutions. Another question found that 33% of respondents have a dedicated digital preservation department while 59% do not.

How Do We Define Digital Preservation?
Out of 20 possible options, 19 were considered in scope for over half of the respondents.

Metadata creation/extraction and preservation planning are considered within the scope of digital preservation for over 90% of the respondents.

Emulation was only considered in scope for 27% of the respondents.

In-house vs Outsourced Functions
Most institutions do all their digital preservation functions in-house. Preservation planning, selection for preservation, development of guidelines for content creators, and development of preservation policies and strategy are done in-house by over 98% of the respondents who are currently doing these functions.

Outsourcing of tasks is infrequent but is used for functions such as digitization, secure storage management, and creation of access copies.

Is What We Have Working Well?
When asked if their current organizational model for digital preservation worked well, respondents who were expecting significant data growth in the next year (=50%) were far more dissatisfied than those with less expected data growth.

The relatively high response rate for the "neither agree or disagree" option may be an indication of respondents' uncertainty about how success in digital preservation is defined and measured.

What We Want

But Not Just Any People: We Need the Right People in the Right Place!
Respondents identified positions whose duties have a direct correlation to content as being the most understaffed. As content increases, the amount of staff also needs to increase. Positions that do not have a direct correlation with the amount of content generally showed a lower desired number of FTEs.

We Need More People!
Most respondents do not think they have the appropriate level of FTE to do digital preservation work successfully. Taking all the positions together, on average institutions have 11.4 FTE doing digital preservation work, but ideally they would like to have 21.14 FTE - a gap of 9.73 FTE.

How Do the Right People Gain the Right Skills?
Retraining is the trend. The large majority (75%) of respondents retained existing staff to perform digital preservation work while 35% said they hired new experienced staff.

What Makes Them the Right Person?
The qualification rated the highest when hiring a digital preservation manager is passion/motivation for digital preservation (58% said this is extremely important; 34.57% said it's very important). The second highest rated qualification is knowledge of digital preservation standards, best practices and tools (57.32% said this is extremely important; 37.8% said it's very important). Surprisingly degrees in LIS, Computer Science and/or digital preservation/curator were not rated as highly.

Page 26