CARL Digital Preservation Working Group

Phase 2 Survey Report on Digital Preservation Capacity and Needs at Canadian Memory Institutions

Grant Hurley and Kathleen Shearer

Survey subgroup members: Lise Brin, Alan Darnell, Corey Davis, Susan Haigh, Grant Hurley, Steve Marks, Michael Moosberger.

Finalized: December 12, 2018

Table of Contents

1. Executive Summary

This report presents the summary results of Phase 2 of the CARL Digital Preservation Working Group survey on digital preservation capacity and needs at Canadian memory institutions. The purpose of the CARL DPWG survey is to provide an updated and comprehensive picture of digital preservation activities in Canada and to help identify existing gaps and outstanding needs at Canadian institutions. While Phase 1 of the survey targeted CARL members, Phase 2 targeted a broad array of Canadian memory institutions. The 25 respondents consisted of 7 academic libraries and archives outside of CARL; 13 government-based archives, libraries, and museums at municipal, provincial and national levels; and 5 community-based or non-profit archives, libraries and museums.

All respondents are undertaking **digitization** activities, and most are **collecting** born-digital materials. The main sources of born-digital materials are private donors and permanent records from respondent organizations themselves. Three organizations are not collecting born digital materials due to a lack of capacity to preserve them.

Organizational commitment to digital preservation is in development among respondents. 32% of respondents have language that expresses a commitment to digital preservation published in a strategic plan, mandate or mission statement. 88% have an individual or group responsible for coordinating digital preservation activities across the organization and 44% have a committee or working group related to digital preservation work. Engagement in external organizations, projects and initiatives relating to digital preservation is mixed: 56% of respondents indicated participation in such groups

The status of formal digital preservation **policies and procedures** is relatively weak among respondents, but progress is building in this area as organizations begin to scope and draft policies. 20% of respondents have an approved digital preservation policy or set of policies. However, a strong cohort of respondents are working to document preservation procedures, strategies and plans. 32% have documented procedures, and another 44% have procedures in draft or development. 68% have digital preservation strategies or plans in place. 56% also indicated they have adopted specific digital preservation-related standards, best practices or guidelines.

Uses of **tools for digital preservation** tasks among respondents are low. 28% of respondents are using tools for digital forensics work and 32% are using tools for preservation processing. 16% are using any one tool for preservation processing in production rather than testing.

While most respondents endeavour to give **access** to digital materials, fewer of them appear to be doing so through consistent methods. 68% of respondents use a web-based platform for access, and others use shared folders or computers on-site.

The transition to preservation-friendly **storage** among all respondents has been slow. While 92% of respondents rely on local network storage as one storage option, 59% on average also depend on CDs/DVDs, hard and flash drives and legacy media (such as floppy disks). Adoption of cloud-based storage services is low: 28% of respondents make use of cloud storage services. 48% use tape and 8% make use of replicated storage networks like LOCKSS.

Most respondents have low **staffing** levels devoted to digital preservation. While 96% of respondents have at least one individual with some responsibilities in the area, the proportion of time devoted to digital preservation is small. 61% of respondents have less than the equivalent of one full-time individual working on digital preservation across all staff listed. 49% of all roles listed by respondents had between 0 and 20% of a person's time given to digital preservation responsibilities. Expectations for expanding staffing are also low: 32% said they intended to expand staffing through new hires or reassignment.

Digital preservation programs are largely **funded** through general budgets, but 56% of respondents also rely on short-term funding sources such as grants or awards to accomplish this work. Outside of reliance on IT departments, few organizations have access to additional resources either elsewhere within their organization or externally.

Based on a scoring method used across quantitative responses in the survey, the following picture emerges **overall**. A small cohort of 16% of respondents are just starting out in developing digital preservation capacity. 68% of respondents have programs that are "in progress," with different components at different stages of development, but main areas of weakness in the tool use/access and staffing/funding areas. Another small cohort of 16% of respondents showed greater strength overall as their programs are maturing towards a capacity to perform preservation work on a routine basis.

2. Introduction

This report continues the work of the CARL Digital Preservation Working Group (DPWG) to survey Canadian institutions on the state of digital preservation activities and needs at their organizations. While the first phase of the survey targeted CARL members, the second phase had a much broader scope: Canadian memory institutions of all types. The purpose of the CARL DPWG survey is to provide an updated and comprehensive picture of digital preservation activities in Canada and to help identify existing gaps and outstanding needs at Canadian institutions. In both phases, the survey gathered data on current policies and practices, uses of tools and infrastructure, organizational and governance issues, and the extent of digital holdings at respondent organizations. The second phase was undertaken to broaden the set of survey responses, identify additional stakeholder groups and areas of activity beyond research-driven academic institutions, and to understand if these stakeholders have similar or different capacities and needs in digital preservation.

CARL undertook Phase 2 of the survey between August and September 2018. Respondents were targeted via listservs hosted by the gallery, library, archives and museum communities in Canada, as well as via representatives on provincial and territorial archives associations. The 25 respondents included 7 academic libraries and archives outside of CARL; 13 government-based archives, libraries, and museums at municipal, provincial and national levels; and 5 community-based or non-profit archives, libraries and museums.

The survey questions for Phase 2 remained largely unchanged from the first phase conducted in October to December 2017, with the exception of a handful of corrections or clarifications. CARL respondents from Phase 1 will be contacted to update any aspects that may have changed since their initial participation, and the results of both phases will be compiled together in a final report to be published in early 2019. The results of both phases were also presented at the @Risk North 2: Digital Collections forum in Montréal on November 9, 2018. It is anticipated that the findings will inform the development of strategies, policies, expertise and resource allocation to enable the community to build capacity and help to ensure that Canada’s valuable digital assets are preserved for future generations. In addition, the survey could serve as a way of benchmarking the progress of digital preservation work in Canada.

The following summary report provides an overview of key areas of the survey. Each section of the report summarizes the data collected and notes the gaps and challenges selected by respondents, alongside additional comments and other areas of concern. The report is concluded by a comparison of strengths and weaknesses across respondents and a summary of issues. Instances where there were trends specific to respondent sectors are noted separately. If no sector-specific information is noted, it should be assumed that the results did not differ substantially between sectors.

3. Scope and Types of Digital Collections

92% of respondents indicated that they are collecting born-digital materials and all are involved in digitization activities. Therefore, all respondent organizations have an interest in preserving these digital assets. The three respondents that noted that they are not collecting born-digital materials were government units: a public library, an archive, and an information management unit, who indicated that they are not accepting these materials because they do not yet have the capacity to preserve them. The respondents who are collecting born-digital materials are doing so from two main sources: private donors, and the organization itself in the form of records and publications. Fewer are collecting from other entities. Faculty and students were unsurprisingly a key content source for the academic respondents, and government materials were a focus for government-based respondents.

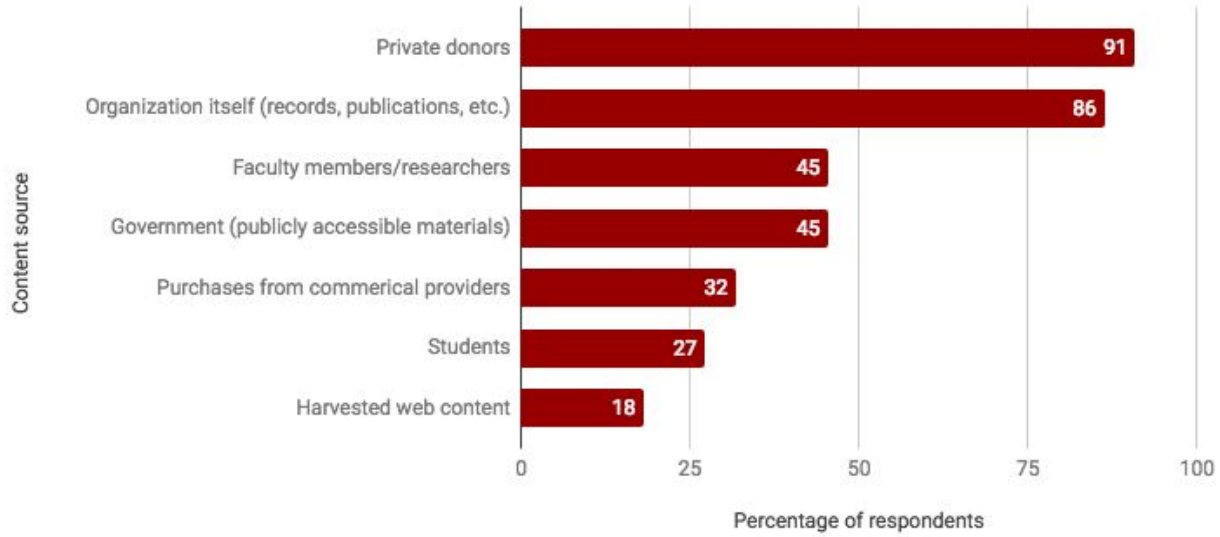

Figure 1: Ranked content sources selected by percentage of respondents. Percentages were calculated out of the 22 respondents who indicated they are collecting born-digital materials.

The **content types collected** show alignment with the content sources. Materials commonly acquired by archives (photographs, moving images, audio, and documents in the form of personal papers and records) ranked the highest. A drop-off follows for material types commonly acquired by libraries, such as purchased publications, databases and datasets. Community/non-profit respondents are primarily collecting just the archival-type materials identified above, which partially accounts for this drop. Research data was selected by 45% of respondents: 5 academic respondents, 1 community/non-profit respondent and 4 government respondents. Web archives followed a similar distribution with 36% of respondents: 3 academic institutions, 1 community/non-profit respondent and 4 government respondents.

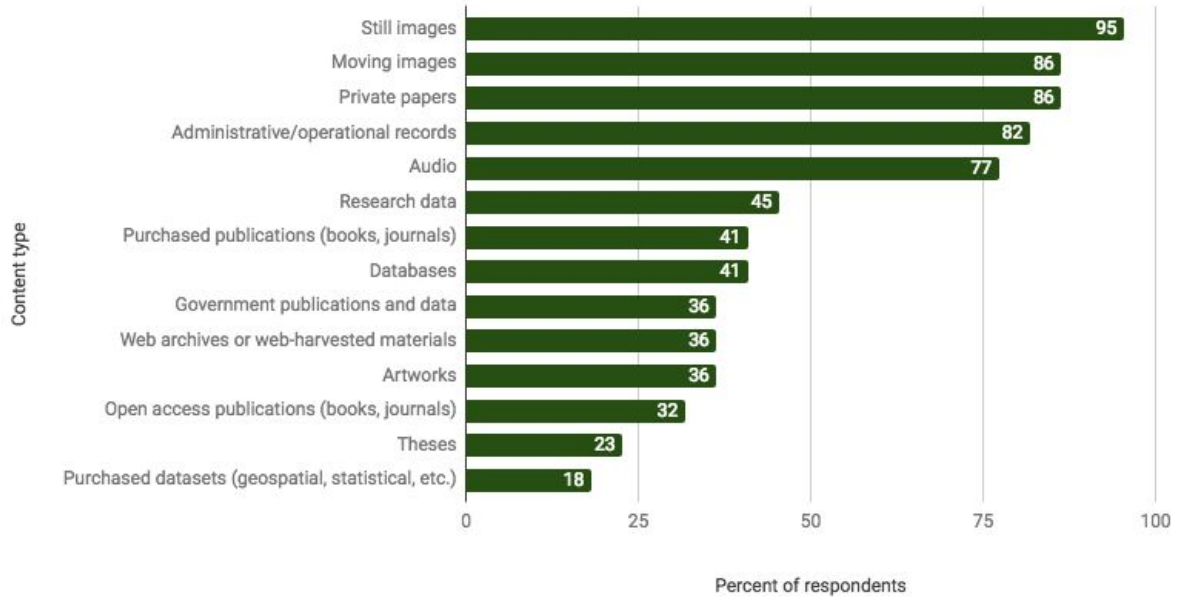

Figure 2: Ranked content types by percentage of respondents. Percentages were calculated out of the 22 respondents who indicated they are collecting born-digital materials.

Respondents ranked the **highest priorities for digitization activities** as photographs, followed by text-based documents and moving images. Some respondents noted in the comments that they do not prioritize digitization based on content type: priorities are determined by other internal needs, such as exhibitions or client requests.

When asked "**What digital assets do you wish you could preserve but currently cannot?**" respondents specified a handful of content types: **university or government-originated administrative records** (4 respondents), **web-based sources** (4, including 2 who specifically noted local news sources as a concern), **3D images** (2), and **e-mails** (2). 5 respondents also noted specific **media carriers**: analogue audiovisual carriers such as film (2) and U-matic tapes (1) as well as specific kinds of digital carriers: floppy disks and digital audio tape (2).

2 respondents felt that their **current preservation capacities were adequate** to preserve the majority of the assets in their collections. However, 2 responded "all of them" to the question "What digital assets do you wish you could preserve but currently cannot?" In other words, they **do not feel they can currently preserve anything** in their collections. 2 others indicated that they are not currently accepting born-digital donations in general, while 1 other respondent noted that "all digitized content" requires preservation.

Key takeaways:

- Most respondents are collecting digital materials and all are pursuing digitization activities.
- While the content sources and types differ, a higher emphasis on records from organizational and private sources points to the need to support archives-specific

workflows for preservation, especially for materials from private donors and records creators within organizations.

- A small proportion of respondents are feeling confident in their abilities to preserve their holdings, while another group is not undertaking activities to preserve assets or are not accepting digital materials due to a lack of preservation capacity.

4. Organizational Commitment and Engagement

One indication of an organization's commitment to digital preservation is specific language in its published strategic plan, mission statement, or mandate related to preservation. **19 respondents (76%) have language that expresses a commitment to digital preservation either published, drafted, or planned**. 6 respondents (24%) have no such language in place. 3 of these latter organizations were community/non-profits, while the other 3 were government-run museums or public libraries.

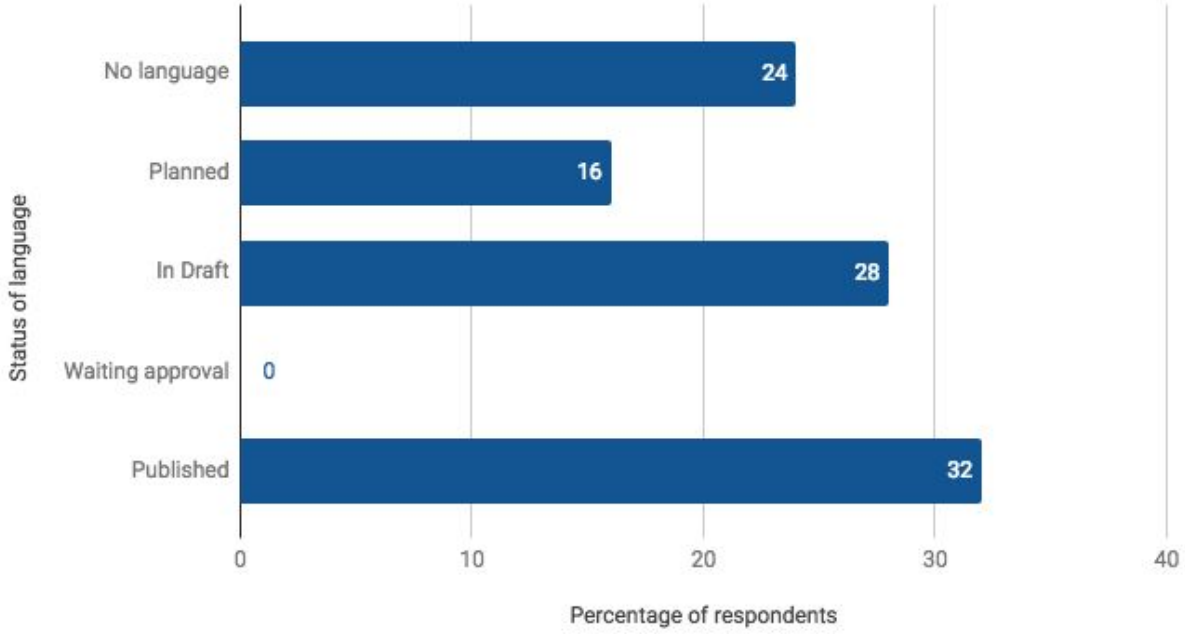

Figure 3: Status of language in a strategic plan or mission/mandate statement expressing commitment to digital preservation by percentage of respondents.

When asked about **areas of activity for digital preservation**, a high concentration (16, or 64%) of respondents noted digital preservation activities are occurring in collections-focused units. 3 (12%) respondents noted digital preservation activities are happening across all units: these organizations were provincial or regional archives. 3 respondents noted corporate records and information management units, 2 (8%) mentioned digitization projects or units, 2 mentioned information technology units, and 2 mentioned digital services or systems units. **13 (52%) respondents have more than one unit or department undertaking digital preservation activities in their organization**.

22 respondents (88%) have an individual or group responsible for coordinating digital preservation activities across the organization. Of these, 9 respondents named senior administrators, such as a university librarian or city archivist, and 8 listed individual archivists or librarians in non-management roles. 5 groups were listed.

Table 1: Examples of titles of individuals or groups responsible for digital preservation

Individuals	**Working Groups or Committees**
● Bibliothécaire en chef de la Bibliothèque ● Chief Records Officer ● City Archivist ● Director of Collections Management ● Director of Access and Conservation Services ● Team lead, Digital Preservation Group ● Digital Initiatives Librarian ● University Archivist ● Volunteer Archivist ● Digital Archivist	● Digital Preservation Group ● Digitization Committee ● Information Management Steering Committee

11 respondents (44%) have a committee or working group related to digital preservation. Of these, 4 respondents named a formalized group whose work is focused on digital preservation, 3 named broader preservation or collections-focused groups, 2 named digital initiatives or infrastructure-related teams, and 2 named an informal group of individuals with interests in this area. Roles identified for these groups included identifying policies and procedures, providing collections oversight, evaluating and implementing hardware and software infrastructure, and linking digital-related strategies with tasks and deliverables. Only 1 of 5 community/non-profit respondents (20%) listed a committee or working group. 4 of 7 (57%) academic respondents and 6 of 13 government respondents (46%) listed a committee or working group.

There was a **mixed level of participation indicated in external organizations, projects and initiatives relating to digital preservation** among respondents. **14 respondents (56%)** indicated participation or membership in regional, national, or international organizations, conferences or projects specific to digital preservation. Engagement varied among sectors. 6 of the 7 academic respondents (86%) responded “yes,” while only 1 of the 5 community/non-profit respondents (20%) said “yes” to the above question. Among government respondents, 7 of 13 responded “yes” (53%).

The following word cloud provides a list of the organizations weighted by the frequency with which they were noted. The highest frequency was 3 times (iPres, the Association of Canadian Archivists, and the Digitization and Digital Preservation Discussion Group) and the lowest was 1 time. 63% of the organizations listed were indicated just once by respondents, which accounts

for their relatively even distribution.

Figure 4: Word cloud of organizations, projects, initiatives or conferences related to digital preservation in which respondents are participants. Words are weighted by the frequency with which they were listed by respondents.

The **rated gaps and challenges related to organizational commitment and engagement** were:

- Lack of resources to pursue organizational change (19 respondents, 76%)
- Lack of resources to participate in collaborative efforts (16, 64%)
- Lack of high-level organizational commitment or support (11, 44%)
- Lack of communication/coordination among stakeholders (8, 32%)
- No gaps/challenges (1, 4%)

Additional gaps and challenges noted were: a lack of business knowledge/understanding among stakeholders at the organization (3 respondents); a lack of staff skills to pursue organizational change (2); the necessity of relying on grants for resources (2); and the high cost of infrastructure and storage (2). In connection with indicating “lack of business knowledge/understanding among stakeholders,” the respondents noted that other units in the organization do not understand the needs or requirements for long-term digital preservation, including storage and infrastructure requirements, or what activities the field entails in general.

One respondent commented: “Our largest challenge is that our IT infrastructure is done through a centralized IT service for the government, who are ultimately responsible for providing business IT services for all government ministries. As our organization plays a unique role in the preservation and long term access of government records, it is challenging for IT to understand the different role and function of archival records for the purpose of digital preservation versus access and use of business records for the function of running a government agency day-to-day.” Gaps and challenges were relatively equally distributed between respondents with

the exception of “lack of high-level organizational commitment or support,” which was indicated by 6 of the 7 academic respondents (86%).

Key takeaways:

- Organizational commitment to digital preservation is in development among respondents.
- Most respondents have a sense of how digital preservation work is to be coordinated and where activities are occurring in their organizations, which provides a necessary base for developing programs going forward.
- Academic and government-based institutions have medium-to-strong levels of commitment to digital preservation as expressed by language in strategic plans, committees and working groups, and engagement with external groups, while community/non-profit groups ranked lower in the latter two measures.
- Respondents indicated higher levels of uncertainty at the organizational level when it comes to resourcing organizational shifts towards digital preservation and the ability to work with external groups and projects.

5. Policies and Procedures

Policies

All respondents indicated that they have, or are interested in, implementing a digital preservation policy at their institution. However, only 9 respondents (36%) have policies approved, under review, or in draft, while 16 (64%) do not.

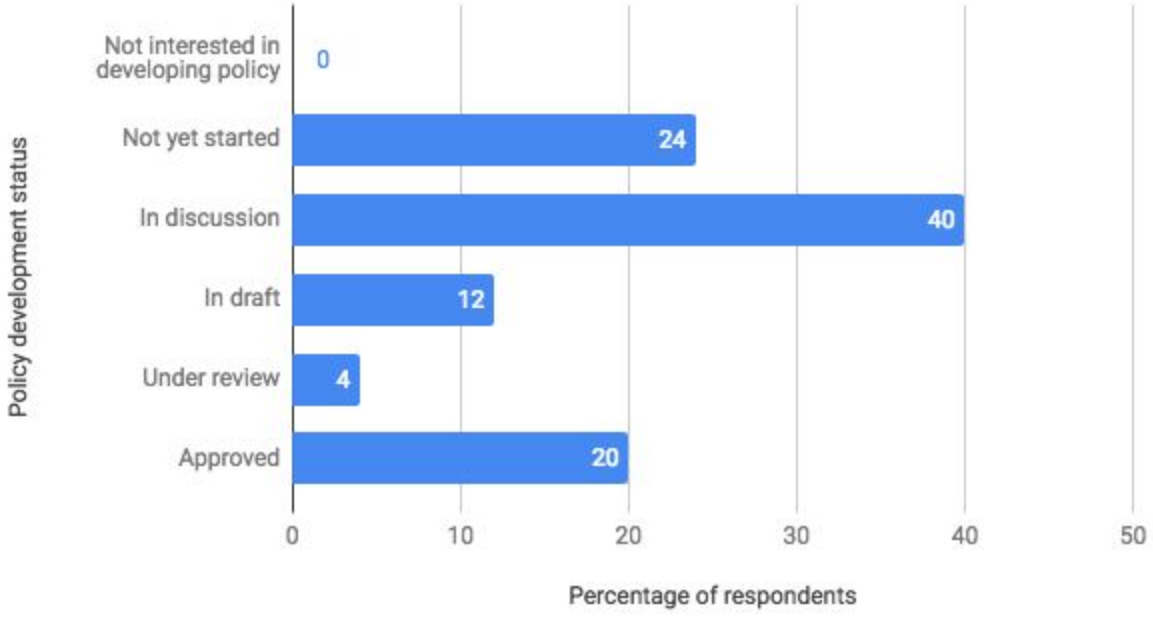

Figure 5: Status of policy development at respondent organizations by percentage of respondents.

Of the 9 respondents who listed a scope for their policies the following topics were noted:

- 4 policies listed treat the mandate or administrative framework under which digital preservation activities occur.
- 3 policies listed treat collections decisions such as acquisition, access, and prioritization for preservation.
- 1 policy listed consisted of individual guidelines for transferring digital archival materials, metadata and digitization.

Table 2: Policy scope examples provided by respondents

"The managed activities the archives engages in to ensure long-term accessibility and usefulness of digital content: digital documentary heritage which is under the control and custody of the archives and digital documentary heritage that the archives seeks to preserve through collaboration and/or partnerships with other agencies, organizations, and governments."
"The policy addresses the standard elements as described in the Digital Preservation Management Workshop."
"Elle définit son cadre juridique et administratif, sa portée ainsi que les champs d'application. Elle couvre les principes et engagements en termes de prestation de service, de préservation efficiente et active des documents numériques."
"The Digital Preservation Policy applies to all digital objects for which [organization] is the primary custodian, including born-digital and digitized material accessioned into the [organization's] collection or Institutional Archives. This material may arrive at [organization] in any format and on any media. Specific preservation decisions are always made in the larger context of [organization]'s institutional priorities and programs, and are contingent on available resources, the needs of [organizations]'s users, and the perceived research value of materials."

The **rated gaps and challenges related to policies** were:

- Lack of time/resources for policy development (21 respondents, 84%)
- Policies are ad-hoc or project specific (8, 32%)
- Lack of knowledge for policy development (5, 20%)
- Policies are not reviewed (5, 20%)
- Policies are not well documented (4, 16%)
- Lack of interest in policy development (3, 12%)
- Policies are not well understood or followed (2, 8%)
- No gaps/challenges (1, 4%)

2 respondents noted that they are awaiting a higher-level organizational shift in order to drive policy development. Other individual respondents noted they are awaiting the implementation of technological infrastructure or do not have staff with experience in policy creation. One respondent commented: "Policies cannot be fully drafted, nor are they implementable until technical infrastructure is in place to facilitate digital preservation work to be completed."

Another noted: “La politique de préservation numérique étant en cours d'adoption par notre organisation, nous n'avons pas encore été à même de constater de lacunes ou défis particuliers outre le manque de temps que nous pouvons y consacrer. [Since our organization is in the process of adopting a digital preservation policy, we have not yet been able to see any particular shortcomings or challenges besides the lack of time we can devote to it.]

Procedures and Workflows

Documented procedures and workflows are not especially well developed among respondent organizations, with most respondents having no procedures, undocumented procedures, or draft procedures only (together, 17 respondents, 68%). However, 8 respondents (32%) indicated they have procedures that are documented.

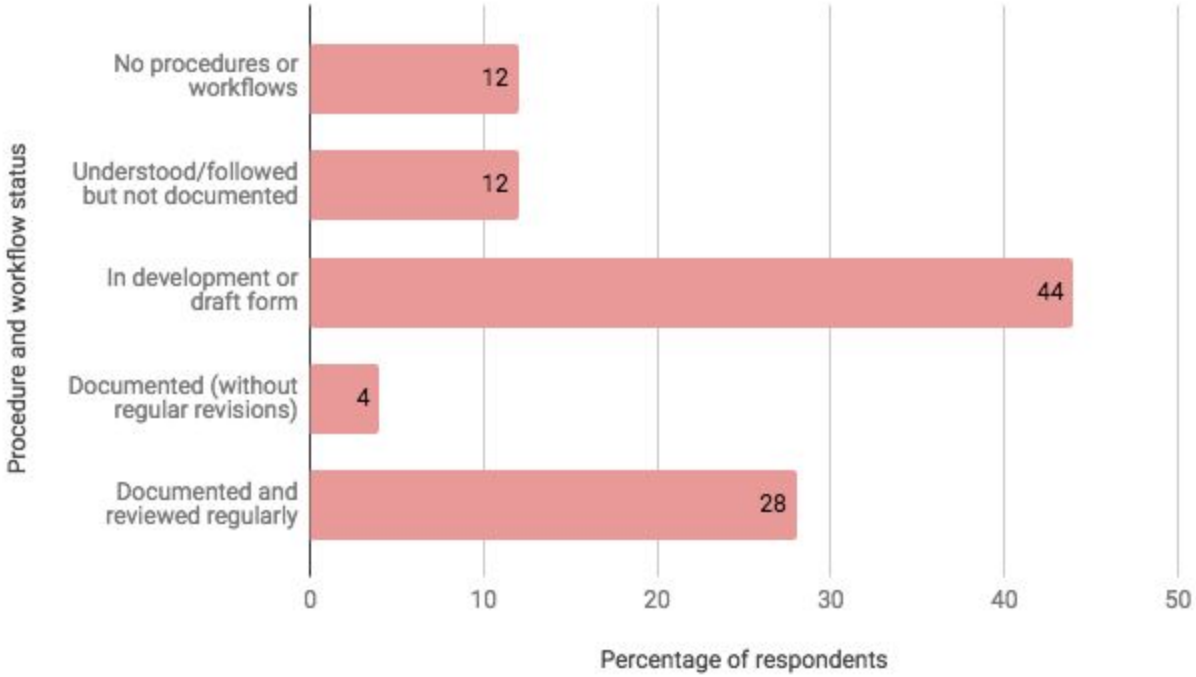

Figure 6: Status of procedures or workflows at respondent organizations by percentage of respondents.

The **rated gaps and challenges related to procedures and workflows** were:

- Lack of time/resources for procedure documentation (19 respondents, 76%)
- Procedures are ad-hoc or project specific (11, 44%)
- Lack of knowledge for procedure documentation (6, 24%)
- Procedures are not reviewed (5, 20%)
- Procedures are not well understood or followed (4, 16%)
- Lack of interest in procedure documentation (3, 12%)
- Procedures are not well documented (3, 12%)
- No gaps/challenges (2, 8%)

Strategies or Plans

17 respondents (68%) have digital preservation plans or strategies in place. Of these, 12 respondents noted that these plans or strategies treat file format standards for preservation and access, including uses within digitization workflows. 3 respondents noted digitization strategies. One respondent each noted the following subjects: prioritization related to Truth and Reconciliation processes, an information management strategy that includes considerations for digital archival retention, the generation of checksums and normalization standards, and a backup policy.

Other Policy-Related Considerations

14 respondents (56%) responded positively to the question, "**Has the institution adopted any particular digital preservation standards, best practices or guidelines?**" Of the 14, the only standards mentioned more than once were OAIS (4 mentions), digitization standards (2) and Dublin Core (2). 86% of academic respondents, 60% of community/non-profit respondents, and 38% of government respondents responded positively to the question.

9 respondents (36%) indicated they had adopted metadata standards for structuring and managing digital preservation metadata. Of the 9, only Dublin Core was mentioned more than once (3 times). PREMIS, METS, PBCore and RAD were mentioned once.

5 respondents (20%) have used digital preservation-related self-audit or gap analysis tools. Of these, 2 indicated they used the TRAC/ISO 16363 standard. Other tools listed were the Digital Preservation Management Gap analysis framework and the AV Preserve Cost of Inaction Calculator. 4 of the 5 positive respondents for this question were government-based organizations at the provincial or national level.

When asked about **interest in pursuing formal ISO 16363 Trustworthy Digital Repository** certification, **3 respondents (12%) indicated that they have undertaken or are interested in undertaking formal certification**. Of the 15 respondents that answered "no," the reasons for not pursuing certification were lack of time/resources (6 respondents), the desire to meet the requirements for certification but not pursue formal certification (3), the use of a third party provider who is or will be certified, and a lack of leadership to pursue certification (2 respondents each). 2 of the positive respondents for this question were government provincial archives; the third was an academic archive. 7 said "I don't know" in response to the question.

Key takeaways:

- The status of formal digital preservation policies is relatively weak among respondents, but progress is building in this area as organizations begin to scope and draft policies.

- A notable presence of documented or drafted procedures, as well as uses of plans or strategies, indicates that positive development in this area is occurring, perhaps motivated by practical concerns such as handling file formats in repositories and undertaking digitization activities.
- Other areas of engagement at the policy level, such as the adoption of standards or use of assessment tools, were relatively low. This circumstance makes sense, as the adoption of standards and other tools for program assessment tends to come as programs gain maturity.
- A lack of resources to develop policies and procedures was identified key challenge by a strong proportion of respondents across the board.

6. Tools and Applications for Preservation Functions

Digital Forensics

The survey included a few questions to gauge interest in digital forensics activities, which enable safer transfer of born-digital materials from legacy media. **28% (7) respondents are using tools for digital forensics**. 6 of the 7 (86%) respondents using such tools were government-based respondents. None of the academic respondents are using these tools and the additional respondent is a community/non-profit. 4 of the 7 respondents use BitCurator, and 2 use both FTK imager and Tableau write-blockers to accomplish this work.

The **rated gaps and challenges related to digital forensics** were**:**

- Lack of staff knowledge/skills (13 respondents, 52%)
- Lack of access to software tools (11, 44%)
- Lack of access to hardware (disk drives, write-blockers, etc.) (9, 36%)
- No gaps/challenges (0, 0%)

Other mentions by individual respondents were that disk imaging is not a priority, or that these processes are still in the planning stage.

Preservation Processing Tools

Respondents showed low use of preservation processing tools: 32% are using such a tool. Of the 8 respondents using tools, 6 are using Archivematica as one of these tools. 3 of these uses are in production and 3 are in testing. Other tools listed were Preservica, Apache Taverna, and Ultima, with one use each. In total, 4 respondents (16%) are using any one preservation processing tool in production. A higher proportion (4 of 7, 57%) of academic respondents are using preservation processing tools versus government respondents (3 of 13, 23%). 1 of 5 community/non-profit respondents is using these tools (20%).

The **rated gaps and challenges rated for preservation processing tools** were:

- Lack of money to support tools (19 respondents, 76%)
- Lack of software/tool support (15, 60%)
- Lack of staff knowledge/skills (13, 52%)
- Lack of access to hardware (9, 36%)
- No gaps/challenges (1, 4%)

In the comments, others noted lack of time (2 respondents) and the fact that the use of these tools is still in the planning stage (2). One museum-based respondent noted: “Museums, as expected, prioritize tools to help manage large physical collections. As such, resources towards digital collection tools is generally a secondary priority. Although this is changing as we collect and create more digital objects, the infrastructure and tools to manage digital collections are not yet on par with those used to [manage] the physical collections.”

Key takeaways:

- Tool use among respondents is low overall, especially when taking into consideration the small number of respondents who are using such tools in production rather than in testing.
- Government-based respondents are focused slightly more on developing forensics capabilities, while preservation processing tools are used more by academic respondents.
- Overall, this circumstance indicates that the majority of respondents are not performing preservation processes on digital materials in their care. The reasons link to similar themes throughout this report: a lack of funding and resources to pursue uses of these tools.

7. Discovery and Access

All but 2 (92%) respondents enable access to digital objects under their care. While the majority - 68% (17 respondents) - use a web-based platform, other transfer methods were also selected, such as a dedicated computer terminal (11, 44%), web transfer or shared folders (10, 40%), and a variety of other methods, such as in response to a direct request (2, 8%), and via email or social media (1 respondent each).

60% of respondents use a digital object repository or discovery platform. These respondents were largely academic institutions (86% of their sector sub-group) and community/non-profit archives (80% of their sector sub-group). In contrast, 45% of government respondents responded “yes” to this question. The types of access systems selected were mixed between respondents. Academic and community-based respondents were the majority users of Access to Memory (AtoM) (7 of the 8 users, or 88%). Of the “other” responses, The Museum System and InMagic were both mentioned twice by different types of respondents. One respondent each listed Preserica, DB/TextWorks, Axiell Emu, and Argus.

Table 3: Access Systems in Use

System	Digital objects only	Descriptions only	Digital Objects and Descriptions
ArchivesSpace			
AtoM		4	4
Blacklight			
ContentDM			
Dataverse			
DSpace			2
Islandora			3
Samvera/Hydra			
Other	1*	1	6

*This product is Preservica, which can have a variety of configurations for access.

The **rated gaps and challenges related to access** were:

- Lack of technological infrastructure (15 respondents, 60%)
- Lack of storage space (13, 52%)
- System/software limitations (12, 48%)
- Privacy/security issues (12, 48%)
- Lack of policies/procedures (11, 44%)
- No gaps/challenges (2, 8%)

Additional concerns expressed by respondents were related to lack of staff time and resources (3 respondents), the “in progress” status of determining needs or selecting a product (2), limited resources to perform backups, intellectual property issues, and the appropriate management of traditional knowledge.

Key takeaways:

- While most respondents endeavour to give access to digital materials, fewer of them appear to be doing so through consistent methods or products, such as web-based repository software, especially in the government sector. One reason may be issues with procurement and adoption of common open source systems favoured by academic institutions by government entities.
- A majority of respondents find that a lack of access to infrastructure and storage space remain barriers to improving access.

8. Storage

Methods

Reliable storage is a key component of a resilient digital preservation program. Respondent organizations are using a variety of storage options, though **most respondents are using local network storage as one option (92%)**. Fewer organizations are using networked storage infrastructures that often provide more reliability, such as cloud networks and tape backups.

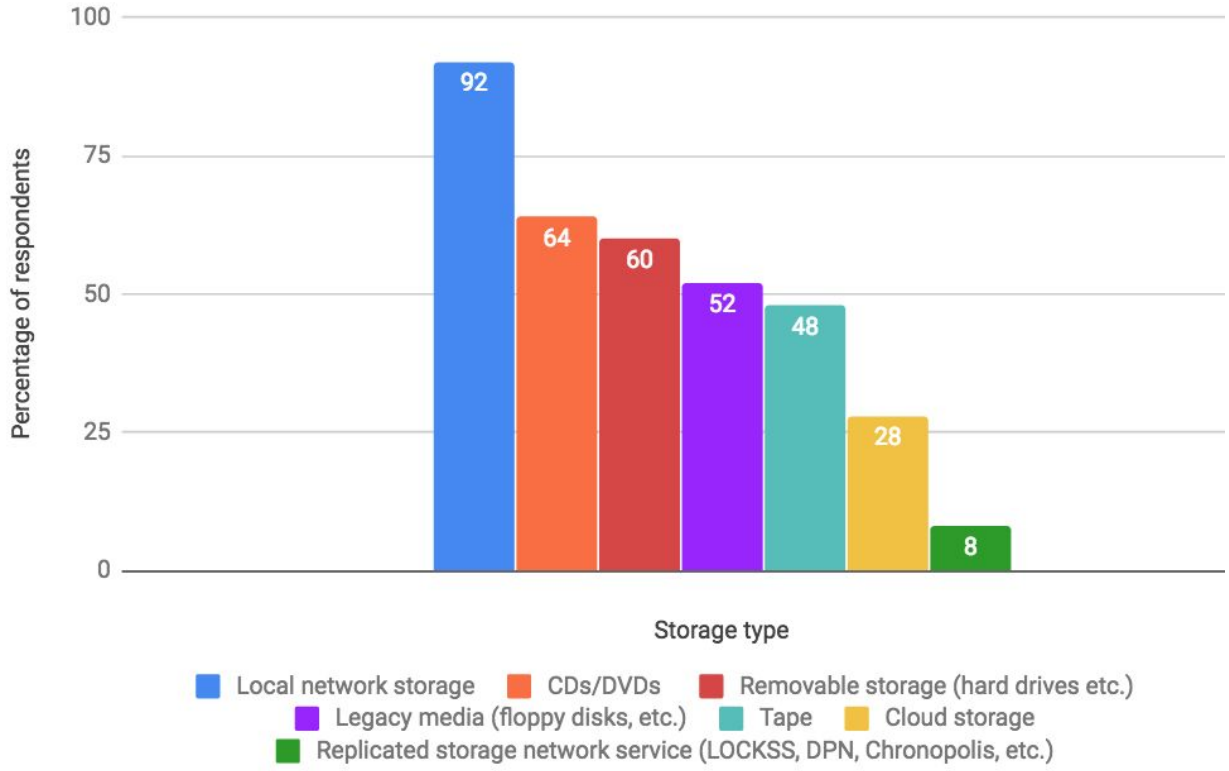

Figure 7: Storage methods selected by percentage of respondents.

Of the 7 respondents (28%) using cloud storage, 2 are using private/community clouds, 2 are using commercial providers, and 2 are using both. 1 respondent did not list the type of service used. Cloud storage was favoured by academic institutions (71% of this group of respondents); 1 community/non-profit and government respondent each also listed cloud storage as an option.

All but two of the respondents are using multiple storage methods and media. The **average number of storage methods** used was 3.48. Academic respondents tended to use more methods (4.14), community-based respondents fewer (2.2 average), and government

respondents met the general average (3.6). Multiple storage places are a good thing when data can be managed well over time, however the widespread reliance on removable hard drives, CDs and DVDs, and legacy media that are prone to degradation and damage, is a potential area of concern. Similarly, reliance on local network storage for preservation can be costly at larger scales and backups and monitoring systems for local storage are required.

Quantities

72% of respondents (18) were able to indicate the amount of digitized content their organization stores. The median amount was 4 TB, with a low of 500 GB and a high of 150 TB. 11 respondents fell between 0 and 10 TB; 4 between 10 and 70 TB; and 3 between 100 and 150 TB. The 3 respondents in this latter category consisted of 2 government respondents and 1 academic respondent.

56% of respondents (14) were able to indicate what quantity of born-digital content their organization stores. The median amount was 1.5 TB, with a low of 50 GB and a high of 130 TB. 12 of the 14 respondents (86%) were storing between 0 and 10 TB of born-digital content. The one respondent with a high proportion of born-digital content (130 TB) was a government-based museum.

Respondents were asked to estimate the **total amount of content stored** if they answered “unknown” for one of the previous questions. 14 responses were given, 5 of which still could not indicate a number (20% of respondents overall). Of the 20 respondents who were able to give two numbers in the prior questions, or a total number in the final question, **the median amount of total data stored was 5 TB**. 12 respondents (48%) had between 10 and 10 TB of content stored overall; 4 had between 10 and 100 TB; and 3 had between 100 and 300 TB; and one outlier listed 3720 TB.

Overall, community/non-profits had the lowest average total storage at 7.15 TB. Academic respondents followed by 29.5 TB on average, and government respondents had the highest amount: 425.55 TB. Discounting the large outlier brings the average for government respondents down to 59.50 TB.

Distributions for Born-Digital Content

60% (15) of respondents were able to estimate the distribution of born-digital data across storage methods. Those who responded indicated on average that **52% of assets were stored on networked systems and 44% on external media**. External media could include hard drives and disks from donors, as well as the use of similar media for storage by the organization itself. A smaller proportion of 6 respondents indicated they had 4% of assets, on average, stored on internal media, such as a personal computer.

Academic institutions and government units tended to have more born-digital assets in networked storage, on average (80% and 60.55%, respectively) while community/non-profit respondents largely have assets on external media storage (93%, with none listing networked storage as holding these materials).

The **rated gaps and challenges related to storage** were:

- High cost of local storage (14 respondents, 56%)
- Lack of local storage (10, 40%)
- Lack of oversight/control over storage (9, 36%)
- Lack of support for storage (7, 28%)
- Procurement barriers (7, 28%)
- Security/privacy barriers (6, 24%)
- Lack of backups (5, 20%)
- No gaps/challenges (4, 16%)

Respondents also noted it can be difficult to estimate needs or costs (2 respondents) and that a lack of local IT resources can be a barrier. One respondent commented, “Our campus ITS is stretched to fulfill their typical responsibilities. We cannot count on them or local resources to provide digital asset storage at our campus. We are too small for this to scale properly.”

Key takeaways:

- Most respondents are using local network storage to keep at least some of their assets, however, there remains widespread reliance on external media and legacy media for storage. Reliance on these media present significant risks where media are not backed up, can be prone to hardware error, may degrade quickly over a short time, or are already obsolescent.
- The transition to preservation-friendly storage among all respondents has been slow, especially for government-based organizations and community non-profits. A better understanding of the barriers to adoption in this area is needed.
- High storage costs remain an issue for over half of the respondents.

9. Staffing

Staffing with responsibilities for digital preservation among respondents is low. While a few respondents have a strong staffing contingent for digital preservation, most respondents have a handful of people for whom digital preservation is a small portion of their responsibilities. Statistics on staff break down into two parts: how many staff members at an organization have responsibilities for digital preservation, and what percentage of their time is devoted to digital preservation-related work.

Number of Roles

The majority of respondents (96%) have 1 or more roles with responsibilities for digital preservation, with 25% (5) having 2 roles, 13% (3) with 5 roles, and 17% (4) with more than 5 roles.

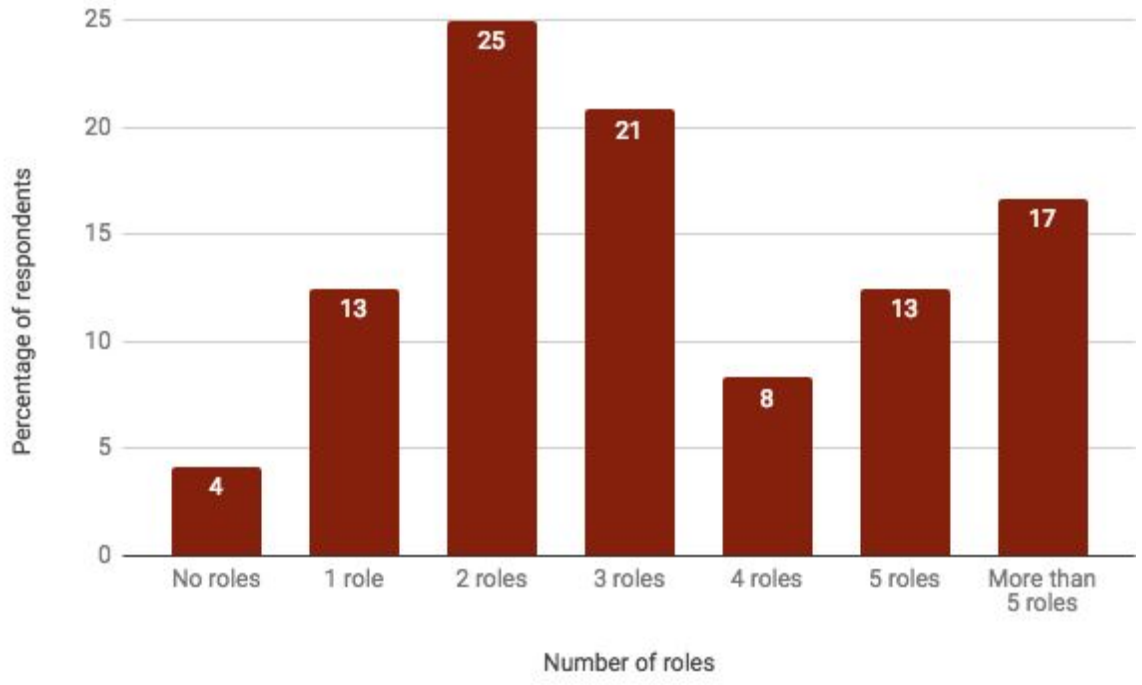

Figure 8: Percentage of respondents by number of roles dedicated to digital preservation. Note: 1 respondent indicated “unknown” for this question and was not counted in the percentage values.

Job titles for digital preservation-related roles varied widely. Of the 70 job titles listed, 17 were archivists, with 5 of these specifically listed as digital archivists. 4 each had “systems” or “technician” in the title; 5 titles listed indicated they are volunteers or students. A strong proportion of titles listed belong to senior or management-level individuals. 15 titles had “senior,” “manager,” “coordinator,” or “director” in their title. In addition, 4 individuals were listed as the heads of their organizations, such as university librarians or city archivists.

FTE Values

Respondent-entered FTE values showed a **low level of staff time directed toward digital preservation work**:

- 4 respondents (17%) that entered FTE values have 1 or more roles that are entirely devoted to digital preservation at 100% FTE.
- 5 more respondents (22%) have roles with digital preservation responsibilities that add up to 100% FTE or more.
- Therefore, 61% of respondents have less than the equivalent of one individual working on digital preservation overall.

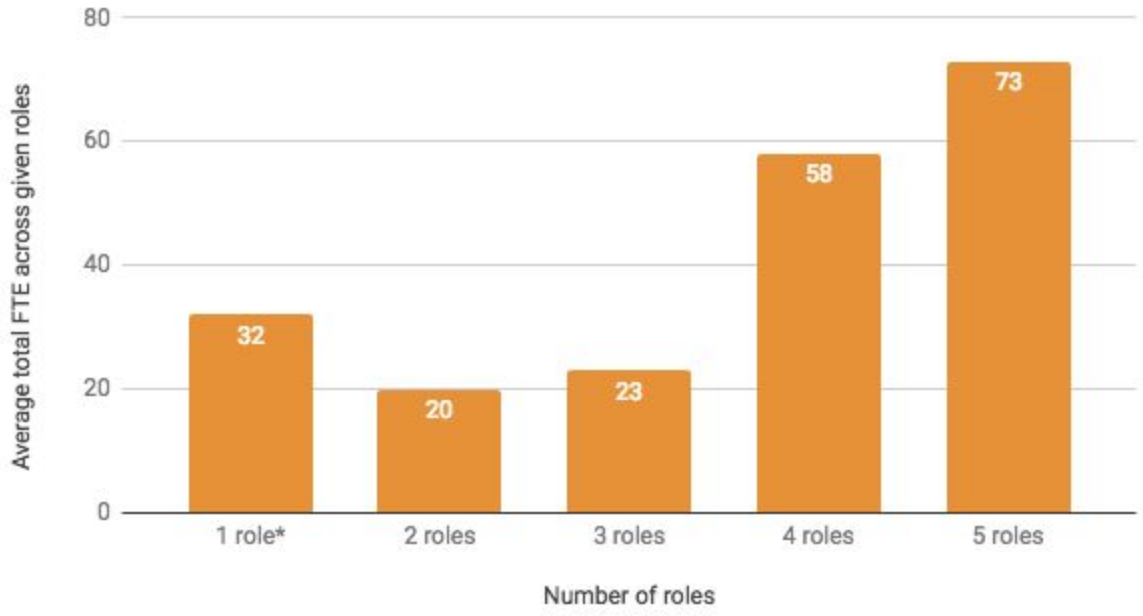

Figure 9: Average total percent FTE based on number of roles at a respondent organization. *Note: Also counted as “one role” in this chart were respondents that listed more than one job title but only one FTE value. Note that two respondents were disregarded as they did not include FTE values.

For respondents with 1-4 roles listed, the average total FTE across these roles was 73% FTE. This average rises to 125% FTE when counting the respondents with 5 roles, including the 2 respondents with 5 full-time roles. However, of the 68 roles with FTE values assigned, 49% (33 roles) were between 0 and 20% FTE. 34% (23 roles) were between 20 to 100% FTE and the remaining 18% (12 roles) were 100% FTE roles.

The following values are specific to individual sectors:

*Academic - 6 of the 7 respondents included**
Total number of roles with digital preservation responsibilities listed: 22
Average number of roles per respondent 3.6
Average total FTE per respondent: 99%
Number of single positions with 100% FTE: 0
Number of institutions with total of 100% FTE or more: 2

Community/non-profit - 5 respondents included
Total number of roles with digital preservation responsibilities listed: 12
Average number of roles per respondent: 2.4
Average total FTE per respondent: 132.9%
Number of single positions with 100% FTE: 6
Number of institutions with total of 100% FTE or more: 2

*Government - 12 of the 13 respondents included***
Total number of roles with digital preservation responsibilities listed: 34
Average number of roles per respondent: 2.83
Average total FTE per respondent: 123.33%
Number of single positions with 100% FTE: 6
Number of institutions with total of 100% FTE or more: 5

*One respondent disregarded because they entered “unknown.”
** One respondent indicated they have 15 staff but did not enter titles or FTE values.

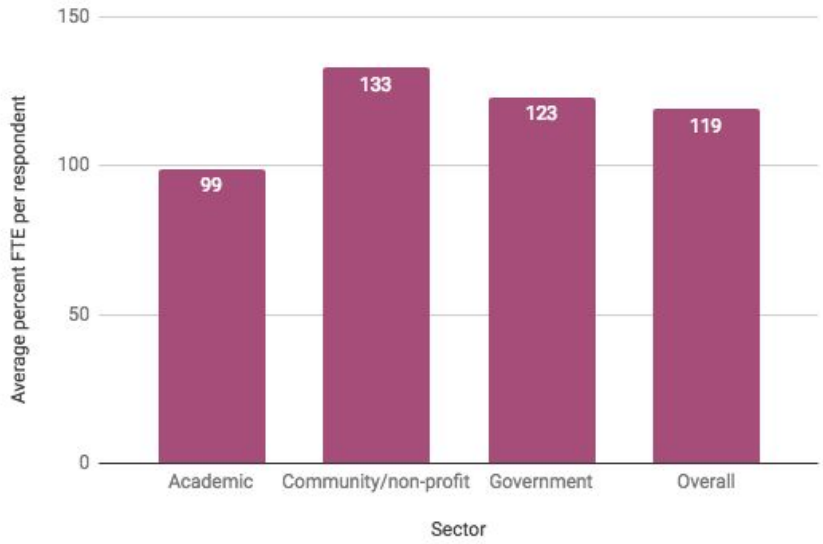

Figure 10: Average total FTE values in percentage per respondent, by sector and overall.

Overall, total FTE for all roles listed was 2739.5; the average total per respondent (including respondents with no roles) is 119.10% FTE, or 1.19 FTE in whole numbers.

Note for the values in this section: the number of roles were counted if a job title was listed, regardless if FTE values were applied or not. FTE values and averages were calculated only against the total number of FTE values given by respondents. Note that 1 respondent responded “unknown” to the staff question; another indicated they had more than 5 staff but did not indicate FTE values.

Expectations for extending staff responsibilities for digital preservation were low. **8 respondents (32%) said they intended to expand staffing**. 2 of these indicated this would happen through new hires; 2 through reassignment; and 4 through both methods. Only 1 academic respondent indicated they would expand staffing, though both methods of staffing were indicated. 3 community/non-profits indicated they would expand staff (2 by reassignment and 1 by new hires) and 4 government respondents indicated they would expand staffing, 3 of whom also indicated both methods.

The **rated gaps and challenges related to staffing** were:

- Lack of funding for new positions (22 respondents, 88%)
- Lack of resources for training/professional development (12, 48%)
- Lack of staff knowledge/skills (9, 36%)
- No gaps/challenges (0, 0%)

No additional gaps or challenges were listed. One respondent commented: “There are currently no FTE positions remaining for the organization therefor[e] any additional staffing will need to be filled through contract positions (budget allowing) as we have done with most other functions for the department.” Another said, “We are a very small staff. It can be a struggle to get work done.”

Key takeaways:

- Most respondents have low staffing levels devoted to digital preservation. While many organizations have at least one or two individuals with some responsibilities in the area, few organizations have the equivalent of a full time individual for digital preservation, and fewer still have any one person with a sole responsibility towards preservation.
- Few organizations expect to expand staffing.
- The major challenge as stated by respondents is a lack of funding for positions.

10. Funding

76% of respondents (19) are funding digital preservation through general budgets. However, the most common second source of funding was grants or awards (14, 56% of respondents) followed by gifts or endowments (9, 36%). 2 respondents listed grants or awards as the sole source of funding.

72% of respondents (18) did not know what percentage of their organization’s budget was dedicated to digital preservation. Of the 7 that responded, 2 said “8%,” 1 said “3%,” and 2 said “less than 1%.” 2 others gave dollar figures: “$5000” and “$98,000,” the latter of which was noted to include student salaries.

While **13 respondents (52%) indicated they expected budget increases in the next 1-2 years**, few of these could indicate what this expected increase would be. 2 respondents indicated 10%, and one indicated 10-20%.

Indication of resources coming from other stakeholders was mixed. 7 respondents (28%) indicated “none.” 3 of these respondents were community/non-profit organizations. By contrast, government-based respondents accounted for the majority of uses of IT departments - 11 of the 14 selected (79%). Only 2 of the academic respondents listed consortia as a source of resources.

Table 4: Sources of external resources for digital preservation activities

Internal departments	**External organizations**
IT department (14, 56%) Larger government unit (1, 4%) Legal department (1, 4%)	Professional association (4, 16%) Regional consortium (3, 12%) Regional partners or stakeholders (2, 8%) National consortium (1, 4%)

The **rated gaps and challenges related to funding** were:

- Funding is not sustained (18 respondents, 72%)
- Allocation of resources is too low in comparison to needs (14, 56%)
- Lack of business plan (7, 28%)
- No gaps/challenges (2, 8%)

“Other” comments noted that funding is awaiting approval or the determination of needs. One respondent commented that “Many projects are driven by one-time funding, but commitments are long-term.” Another noted “As we are still determining our overall digital needs, we have yet to determine what the overall cost is that we require to go forward. Once that is completed, we will need to obtain additional funding and then determine how we can go forward with our requirements.” The respondents who noted “no gaps/challenges” were both municipal archives.

Key takeaways:

- While general budgets are a key source of funding, the reliance on short-term funding such as grants and awards for a long-term activity like digital preservation is cause for concern.
- Most respondents had difficulty in estimating how much money is directed to digital preservation work.
- Increases in funding over the next 1-2 years are expected by just over half of the respondents.
- Outside of academic-based respondents who indicated membership in library consortia, many respondents do not have access to additional resources inside or outside of the organization.
- Most respondents from all sectors do not feel that funding resources are adequate.
- Additional funding challenges relate to difficulties in estimating costs and advocating for needs based on unknown costs.

11. Organizational Maturity

Respondents were asked to **rank the status of their digital preservation programs in general terms** using a standard maturity scale, from 0 (No activity) through to 5 (Optimized - processes are mature and continually improved). 1 respondent (4%) felt that they were at 0 - no activity. 7 (28%) felt they were at the initial stage, and 9 (36%) felt they were one step up at the repeatable stage. 3 (12%) felt they have reached a defined stage, and 5 respondents (20%) felt they are at the managed stage. No organization indicated that they were at the optimized stage.

64% of respondents placed themselves at stages 1 or 2, 32% at 3 or 4, and none at stage 5. The average maturity score overall was 2.16. Average scores by sector did not differ significantly.

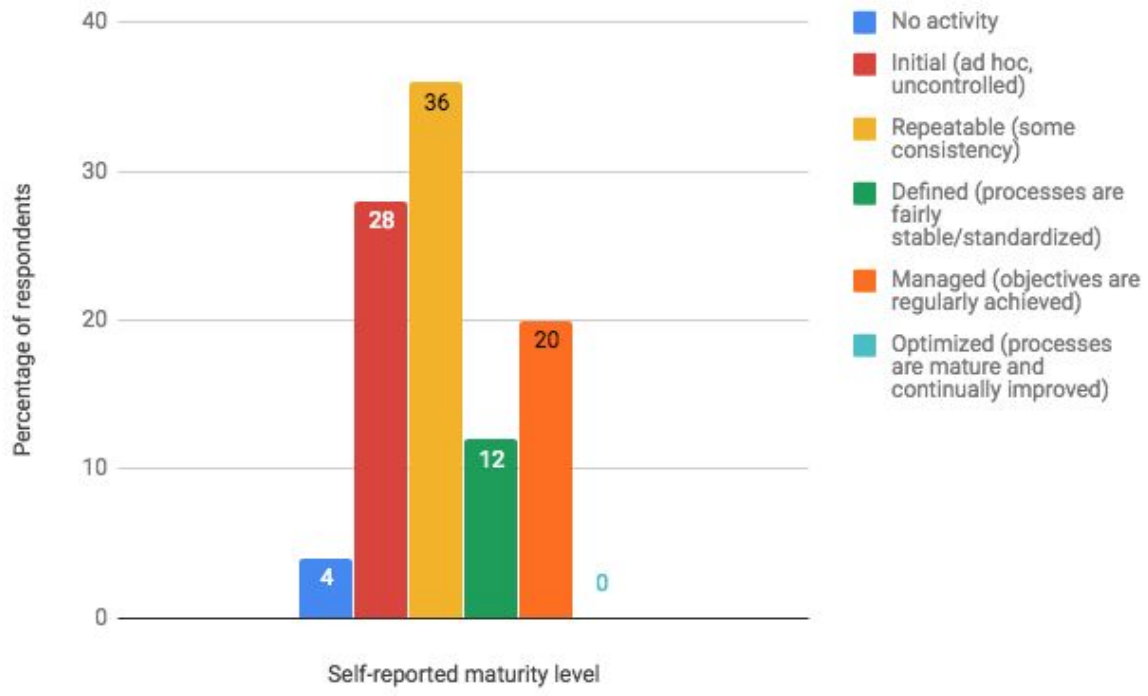

Figure 11: Self-reported maturity scales by percentage of respondents.

Respondents finished the survey with the opportunity to add additional comments. When asked "**Are other organizational, policy, technological, and resource issues preventing digital preservation capacity?**" they noted the following:

Table 5: Comments in response to "Are other organizational, policy, technological, and resource issues preventing digital preservation capacity?"

"We are a very small university with limited capacity to implement a digital preservation strategy. This is our main challenge."
"Lack of a dedicated policy office, lack of a dedicated individual for intellectual property, lack of a digital preservation policy, lack of leadership in these areas at a high level, lack of resources needed for digitization."
"Oui. Les restrictions budgétaires des dernières années viennent entraver la capacité à maintenir l'intégrité et le développement de la collection numérique. Ce sont principalement les ressources qui ont été affectées. Sans financement adéquat pour assurer la conservation pérenne du patrimoine documentaire ..., c'est toute la mémoire collective qui est à risque."

> “2 main ones:
> 1) funding for systems & storage (even open-source systems need to run on something)
> 2) staff who can devote time to both the technical aspects & the ingest / cataloging etc archival practice of digital preservation. So for example, we have discussed setting up AtoM here but would need to devote IT dev resources to standing it up; server resources for hosting it; storage resources for content; and library staff resources to ingest, catalog etc -- none of which is currently easily available.”
>
> “Finding knowledgeable staff is a challenge continuously.”
>
> “No money and low interest in increasing staff time to do digital preservation work.”

Key takeaways:

- The majority of respondents (68%) do not yet see a defined or formalized set of digital preservation activities as core to their operations, while a smaller group (32%) believe they are starting to move towards maturity with more clearly defined processes and the meeting of objectives.

12. Strengths and Weaknesses

As an exploratory measure, respondents were scored across key areas of the survey that contained quantitative information: organizational commitment, policies and procedures, tools and technologies for preservation and access, storage, and staffing and funding. Appendix 1 contains the scoring rubric. The intention was not to assess respondents against a preexisting standard, but to mark areas of strengths and weaknesses to provide for a method of comparison. The method used was to assign points for areas that showed investment or progress and create a simple score out of three for each section. A total of fifteen points were added together from five sections for a final score. The scores are meant to express relative capacity within the context of the responses and therefore should not be taken as absolute values. For example, a score of 14 out of 15 points indicates high capacity within the cohort of respondents, but not absolute potential capacity for that organization. Similarly, it is not possible to measure a digital preservation program as having this or that percentage of content “preserved” in the context of a survey, as understandings of what functions this requires will vary based on the needs and resources at hand. Different content types will require different strategies and resources, and preservation approaches are always in flux as technologies change and materials age. Rather, digital preservation programs are best evaluated based on the sets of functions and activities they are able to support relative to their specific needs.

The lowest score was 3.5 and the highest was 13; the average score overall was 8.58. The average scores per sector were similar:

- Academic: 9.43
- Community/non-profit: 8
- Government: 8.5

The following chart shows the distribution of scores.

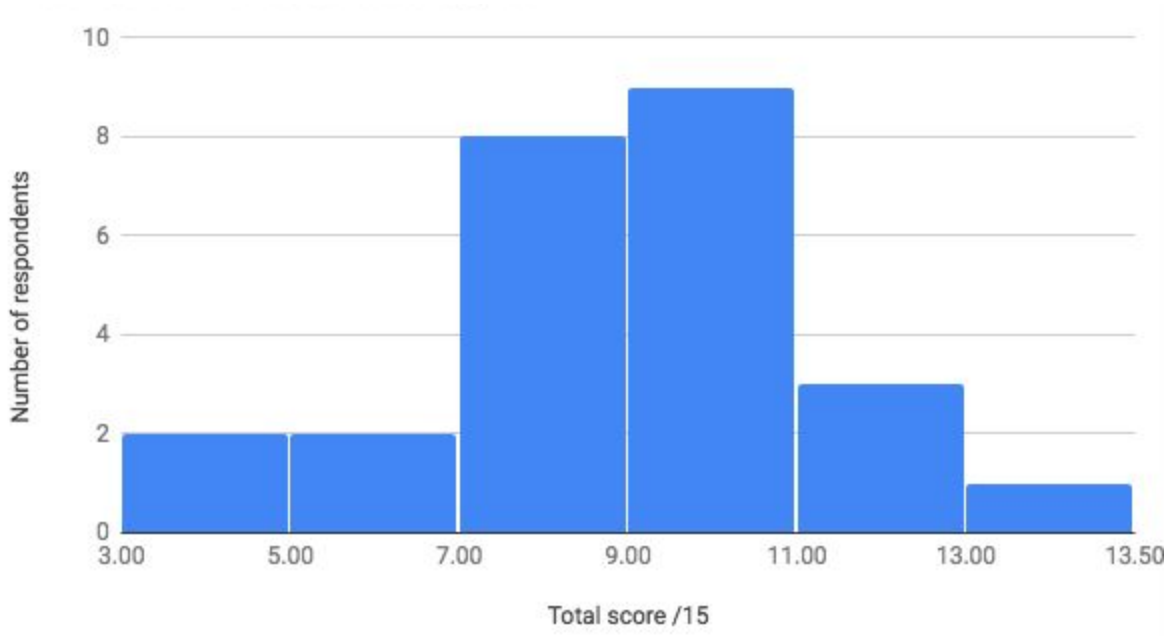

Figure 12: Distribution of scores out of 15 by number of respondents.

Table 6: Score distribution against relative capacity out of 100.

Score range	% of respondents	% relative capacity
0-3	0	0-20%
3-5	8%	20-33%
5-7	8%	33-46%
7-9	32%	46-60%
9-11	36%	60-73%
11-13	12%	73-86%
13-15	4%	86-100%

Insufficient data gathered from respondents in questions regarding total FTE values and total materials expenditures meant that the relationships between scores and these values could not be computed, as was done in Phase 1. Somewhat more consistent data was gathered for overall budgets, and the related chart is presented below against total scores is below. The key message is that there is no clear relationship between size of overall budget and total score. With the exception of one organization with a very small budget, average scores remained about the same (and even lowered slightly in some cases) across budget ranges.

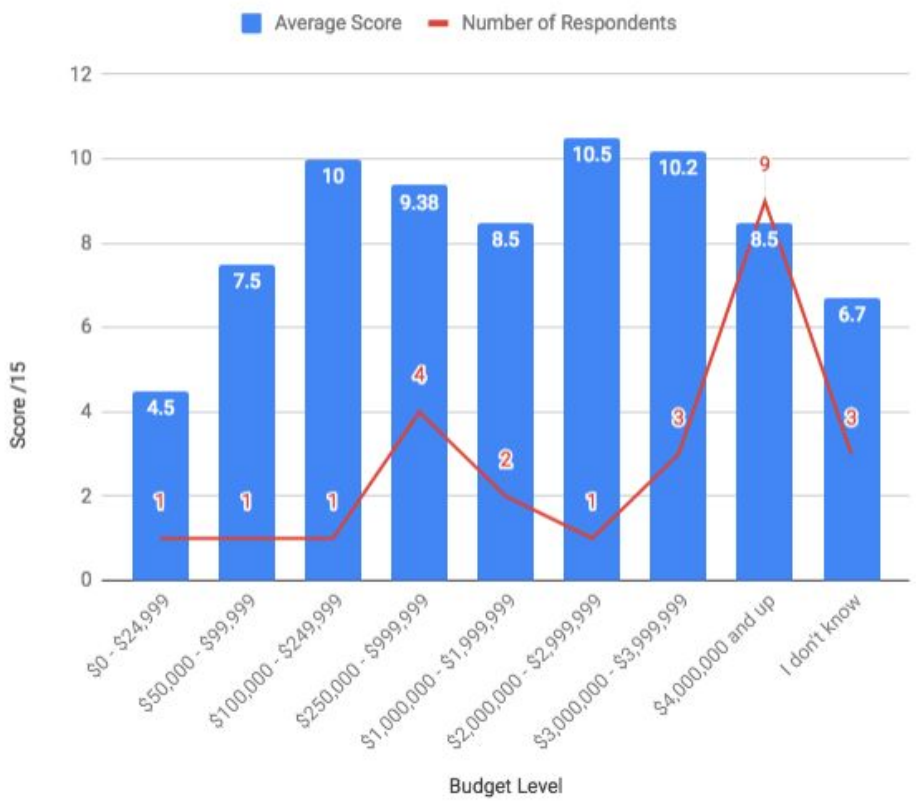

Figure 13: Budget range against average total score for range out of 15. The red line indicates the number of respondents per budget category.

Scores showed similar trends across the board that are in line with the general results. The large proportion of respondents fell into a middle cohort of scores, with smaller groups of low and high scoring respondents on either side:

- A small group of 4 respondents (16%) scored relatively low, between 3 and 7, and represent a group just starting out at about 20-40% relative capacity. These respondents tended to be weaker across all functional areas. These respondents consisted of two public libraries and two community/non-profit organizations.
- A large middle cohort of 17 scorers between 7 and 11 (68%) showed relative capacity as ‘in progress’ but improving at about 40-70% relative capacity. This group consisted of 6 academic respondents, 2 community/non-profit groups, and 9 government-based respondents. The weakest areas of activity among this cohort were tool use/access and staffing/funding.
- 4 respondents (16%) scored 11 and above and showed growing strength in capacity, with 70-90% capacity overall (when ranked against fellow respondents, rather than against an absolute value for capacity). These respondents tended to be stronger across all functional areas, though they had slightly lower scores in the tool use/access and staffing/funding areas. The group consisted of two government-based archives at the provincial level, a community/non-profit organization, and an academic institution.

When mapped to the maturity areas in the “Concluding Comments” section, there is a general observable correlation, though more respondents put themselves at the higher “defined” or “managed” levels than the scores show.

13. Conclusion

As a group, the respondents in Phase 2 of the survey are varied both in terms of size and sector, but their activities, challenges and concerns coalesced around the central problem of doing the work required to operate a digital preservation program across a number of functional areas. While most are collecting born-digital materials, and all are digitizing physical assets, most are not using tools to perform digital forensics or preservation processing, and few have individuals tasked with performing digital preservation-related work on a full-time or full-time equivalent basis. Like CARL respondents, Phase 2 respondents are working towards improving organizational commitment overall, and finding progress in the area of policy development by starting with procedures and working towards formalized policies. They feel generally more confident with storage, though uptake of cloud-based or other digital preservation-friendly storage among Phase 2 respondents is low, and there is a significant amount of reliance on riskier media for storing digital assets. These include legacy media that needs safe migration, and risky uses of hard drives and other external media that are not designed for long-term storage. There was less than full support for web-based access methods, which reflects the varied nature of the organizations in question, but overall respondents seemed fairly confident in this area.

A notable area where respondents differed from Phase 1 was in their interactions with external groups or projects concerned with digital preservation, and uses of additional resources apart from IT departments. Markers for both were low in contrast to CARL respondents; outside of the frameworks for collaboration and resource-sharing within the academic community, it appears that other stakeholders in digital preservation are largely going it alone, possibly due to resource constraints, policy barriers, or the lack of a culture of collaboration more generally.

Overall, there is a clear sense in the comments provided by respondents of a difficult Catch-22 when it comes to building digital preservation programs. Costs for digital preservation are unknown or difficult to estimate without performing more sustained preliminary work, and current resources are not sufficient to move this preliminary work forward. As a result, it makes it difficult to build a business case, reallocate resources, or advocate for more funding, especially for institutions who are not willing to take risks when costs are unknown. These issues are summarized in one comment from a government-based museum: “Our biggest challenge is the costs associated and fear of these costs, staff time to concentrate on preservation activities, and Canadian only consortium to participate in due to government data storage regulations which prevent our participation with more established partnership groups.” Size of organization seems to bear little relation to capacity. Larger budgets do not translate to increased capacity, and most respondents are struggling with resource issues in relation to needs. That said, as in

Phase 1, there is a minority group of organizations in different sectors who have found success in building different components of a digital preservation program at the same time, and therefore have higher relative capacity overall. A more in-depth understanding of their organizational cultures and steps towards success would lend greater depth to the survey results.

The results of Phases 1 and 2 of the CARL survey provide a detailed picture of the capacity for Canadian memory institutions to do digital preservation work. The survey represents an important first step. With this information in hand, a discussion of what is needed to support a more cohesive, comprehensive, and coordinated approach to digital preservation is the next one.

14. Acknowledgements

First and foremost, the authors would like to thank the respondents who took important time out of their day-to-day work to gather information for the survey. There would be little to report without you! Many thanks go to Lise Brin and Julie Morin at CARL for their tireless assistance and support in administering the survey. The authors also thank members of the DPWG for feedback on this report, and for their support of the survey initiative overall.

Appendix 1. Scoring Rubric

Organization & Governance Section

1) What is the current state of your organization's commitment to digital preservation, as expressed through language or wording in a strategic plan or mission statement?

No language in strategic plan or mission statement - 0 points
Adding language is planned - 0.5
Updated language is being drafted - 1
Language is in place, but awaiting approval - 1
Language is published and available - 1

3) Is an individual or group responsible for coordinating these activities? **Or** 4) Does your organization have a committee or working group responsible for digital preservation or an aspect of digital preservation activities (e.g. policy)?

"Yes" in either column counts as 1 point; "No" or "I don't know" as 0.

8) Is your organization a member of, or participant in, any regional, national, or international organizations, conferences or projects specific to digital preservation?

"Yes" counts as 1 point; "No" or "I don't know" as 0.

Policies & Procedures Section

1) What is the status of your organization's digital preservation policy?

Not interested in developing policy - 0 points
Not yet started - 0
In discussion - 0.5
In draft - 1
Under review - 1
Approved - 1

3) What is the status of digital preservation procedures or workflows?

No procedures or workflows - 0 points
Understood/followed but not documented - 0.5
In development or draft form - 1
Documented (without regular revisions) - 1
Documented and reviewed regularly - 1

4) Has the institution adopted any particular digital preservation standards, best practices or guidelines?
Or 6) Has the organization used any digital preservation-related self-audit or gap analysis tools?

"Yes" in either column counts as 1 point; "No" or "I don't know" as 0.

Access/Tools for Preservation Sections

1) How does your organization give access to digital materials?

If 'Web platform/repository' selected, 1 point.

5) Is your organization creating forensic disk images?

"Yes" counts as 1 point; "No" or "I don't know" as 0.

8) Does your organization use digital preservation-related processing tools (e.g. Archivematica, Arkivum, Preservica)?

"Yes" counts as 1 point; "No" or "I don't know" as 0; use counts if in testing or production.

Storage Section

12) In what storage systems/media are digital assets currently kept? (Select all that apply.)

If at least two managed storage methods selected ("cloud storage," "replicated storage network service," or "tape") - 1 point.

Under holdings and activities section:

6) Approximately how many terabytes of digitized content has your organization created? (Indicate "unknown" if you are unable to ascertain this figure.) **or** 7) Approximately how many terabytes of born-digital content has your organization collected? (Indicate "unknown" if you are unable to ascertain this figure.)

If were able to indicate a figure for at least one column - 1 point.

8) Approximately what percentage of born-digital content is stored on each of the following types of storage: (Indicate "unknown" if you are unable to ascertain this figure.)

If were able to indicate a percentage figure adding to 100 - 1 point.

Staffing and Funding

1) List all roles that are responsible/accountable for day-to-day digital preservation activities in your organization and indicate what percentage FTE of each is devoted to digital preservation. (Please estimate to the best of your ability.)

Have at least 100% FTE in total staff listed - 1 point.

2) Is your organization intending to expand staff responsibilities for digital preservation?

“Yes” counts as 1 point; “No” or “I don’t know” as 0.

6) Approximately what percentage of your organization’s last completed fiscal year expenditures were dedicated to digital preservation (including salaries, storage costs, tools and technologies, etc.)? (Indicate "unknown" if you are unable to ascertain this figure.) **or** 7) Do you anticipate increasing expenditures dedicated to digital preservation in the next 1-2 years? - By what anticipated percentage? (Indicate "unknown" if you are unable to ascertain this figure.)

Either indicated percentage value in question 6 or were able to indicate anticipated percentage increase in question 7 - 1 point.