CARL Digital Preservation Working Group

Phase 1 Survey Report on Digital Preservation Capacity and Needs Among CARL Members

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 1 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. Phase 1 of the survey targeted CARL members. 26 respondents filled the survey.

All respondents are undertaking **digitization** activities, and all are **collecting** born-digital materials. The main sources of born-digital materials are faculty and students, permanent records from respondent universities themselves and materials from private donors. Web-harvested materials also ranked high.

Organizational commitment to digital preservation is in development among respondents. 38% of respondents have language that expresses a commitment to digital preservation published in a strategic plan, mandate or mission statement. 65% have an individual or group responsible for coordinating digital preservation activities across the organization and 50% have a committee or working group related to digital preservation work. Engagement in external organizations, projects and initiatives relating to digital preservation is strong: 85% 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 and look towards adopting standards. 15% 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. 16% have documented procedures, and another 50% have procedures in draft or development. 57% have digital preservation strategies or plans in place. 50% 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 but growing. 27% of respondents are using tools for digital forensics work and 50% are using tools for preservation processing. 23% are using any one tool for preservation processing in production rather than testing.

All respondents endeavour to give **access** to digital materials, and all use web-based platforms to do so. A variety of repository/access software packages are in use. The most popular are DSpace, Dataverse and AtoM.

The transition to preservation-friendly **storage** among all respondents has been slow. While 96% of respondents rely on local network storage as one storage option, 60% on average also

depend on CDs/DVDs, hard and flash drives and legacy media (such as floppy disks). Adoption of preservation-friendly storage services is strong: 65% of respondents make use of cloud storage services, 62% make use of replicated storage networks like LOCKSS and 58% use tape.

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. 65% of respondents have less than the equivalent of one full-time individual working on digital preservation across all staff listed. 60% 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 high: 62% said they intended to expand staffing through new hires or staff reassignment.

Digital preservation programs are largely **funded** through general budgets, though 85% of respondents could not indicate what percentage of their organization's budget is dedicated to digital preservation. 77% of respondents expect funding to increase in the next 1-2 years, though few could indicate what level of increase this might be. Most respondents rely on additional resources from IT units internal to their organization as well as external resources from regional library consortia.

Based on a scoring method used across quantitative responses in the survey, the following picture emerges **overall**. A little over one third (34.5%) of respondents are just starting out in developing digital preservation capacity. Another third (34.5%) of respondents have programs that are "in progress," with different components at different stages of development, but missing core pieces, such as organizational commitment or policy development. A final third (31%) of respondents showed greater strength overall as their programs are maturing towards a capacity to perform preservation work on a routine basis.

2. Introduction

The digital preservation landscape is complex and rapidly evolving, and memory institutions are increasingly looking for solutions to address the substantial challenges associated with their emerging digital stewardship responsibilities. As part of the first phase of research on digital preservation capacity and needs in Canada, the CARL Digital Preservation Working Group (DPWG) sought input from CARL members on the state of digital preservation activities in their organizations. 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. 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 CARL member libraries. 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.

CARL undertook Phase 1 of the survey between October and December 2017. The respondents consisted of 25 university libraries plus Library and Archives Canada. 14 respondents indicated that they were responding as a library; 10 responded as a library and archives; 1 responded as a library, archives, and museum; and 1 responded as a library, archives, and gallery.

A second phase of the survey targeting non-CARL members, including organizations outside of the academic sector, was conducted August to September 2018 and a companion report has been prepared. The results of both phases were presented at the @Risk North 2: Digital Collections forum in Montréal on November 9, 2018. 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 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.

3. Scope and Types of Digital Collections

All respondents indicated that they are **collecting born-digital materials and all respondents are involved in digitization activities**. Therefore, all respondent organizations have an interest in preserving these digital assets. The survey showed that CARL members are collecting a variety of digital materials from many different sources, but in a group largely consisting of

academic libraries, it is not surprising to find that the most common sources for digital collections are faculty members or students (such as theses and articles in institutional repositories), as well as organizational assets, such as university records. Private donations also scored high, likely due to the large number of collections in digital formats coming into archives and special collections units from community members and organizations wishing to deposit their papers. 17 respondents (65%) are also actively involved in web archiving.
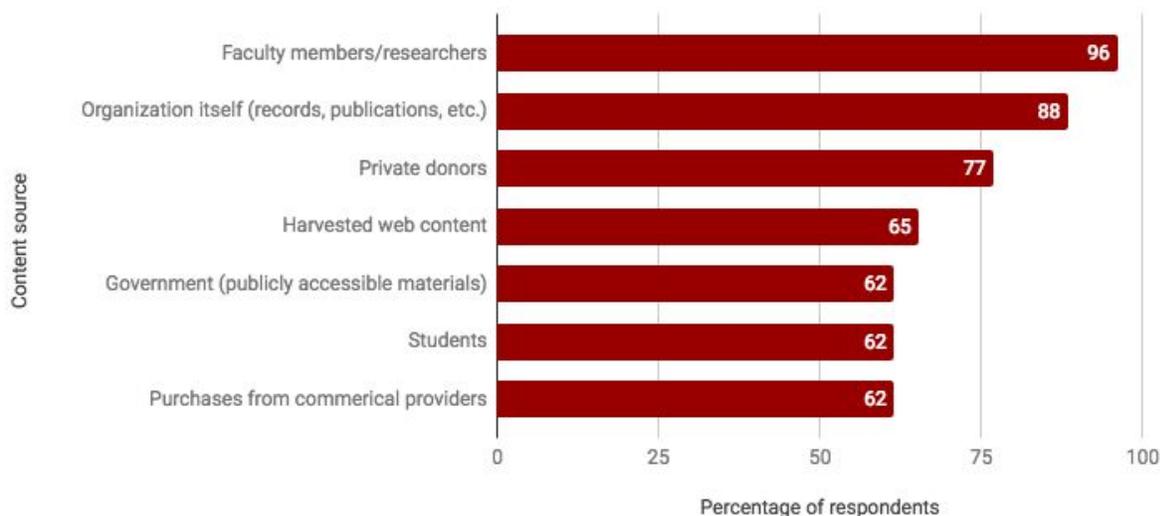
Figure 1: Ranked content sources selected by percentage of respondents.

The **content types collected** are fairly consistent across most respondents with all collecting theses, and nearly all involved in collecting publications, audiovisual materials, structured datasets and research data. The variety of content types being collected by respondents indicates a common need for resources to preserve many different content types and formats.

Figure 2: Ranked content types by percentage of respondents.

Respondents ranked the **highest priorities for digitization activities** as a tie between text-based materials and photographs fol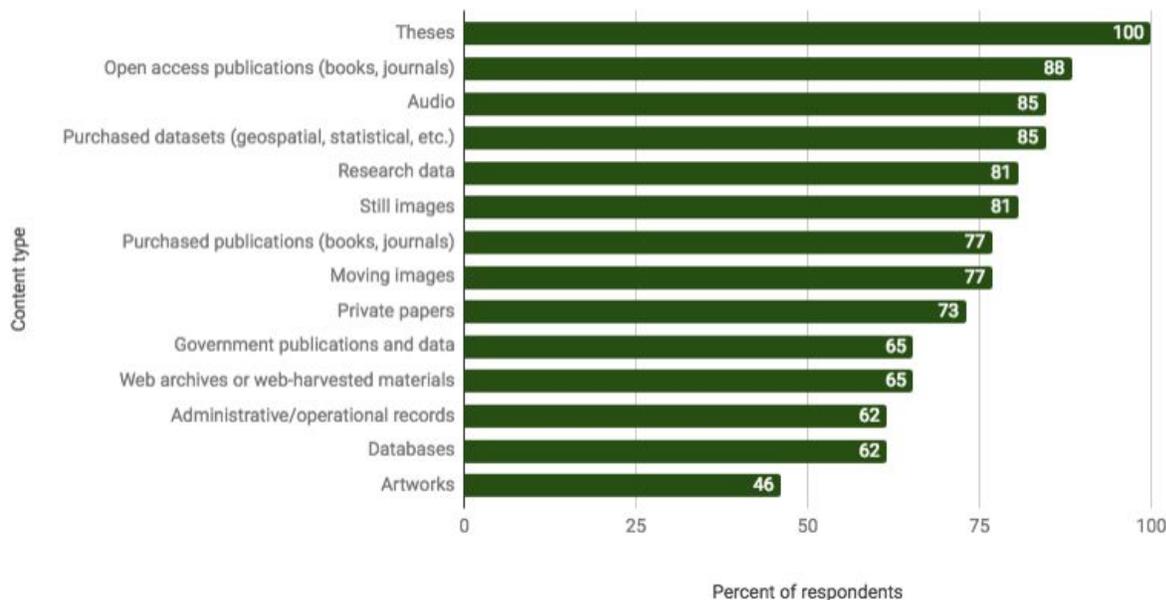lowed by moving images and audio. Some respondents noted in the comments that they do not prioritize digitization based on content type but rather subject area or other considerations, such as research value.

When asked "**What digital assets do you wish you could preserve but currently cannot?**" two respondents noted that the large volume of materials needing to be preserved is the most pressing problem, rather than a specific area or format. 7 respondents indicated that **research data or researcher-produced data in complex, or multi-faceted formats** such as websites, can prove challenging. 6 mentioned **audiovisual materials**, 4 noted **born-digital university records (including e-mail)**, and 3 identified **software and interactive environments** as problematic. One respondent noted "We don't do any preservation at the moment, only back ups, so... all of it!" and another said "No particular type of asset, related more to budget than ability."

Key takeaways:
- All respondents are collecting digital materials from a variety of sources and in a diversity of formats. Resources, training and support targeted at specific content types may be needed.
- Respondents are struggling with the sheer volume of data to be preserved, as well as the complexity of preserving many different format types or materials with complex links or structures.
- Individual respondents are struggling with preserving the digital assets under their care without adequate resources to do so.

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. **20 respondents (77%) have language that expresses a commitment to digital preservation either published, waiting approval, drafted, or planned**. 6 respondents (23%) have no such language in place.
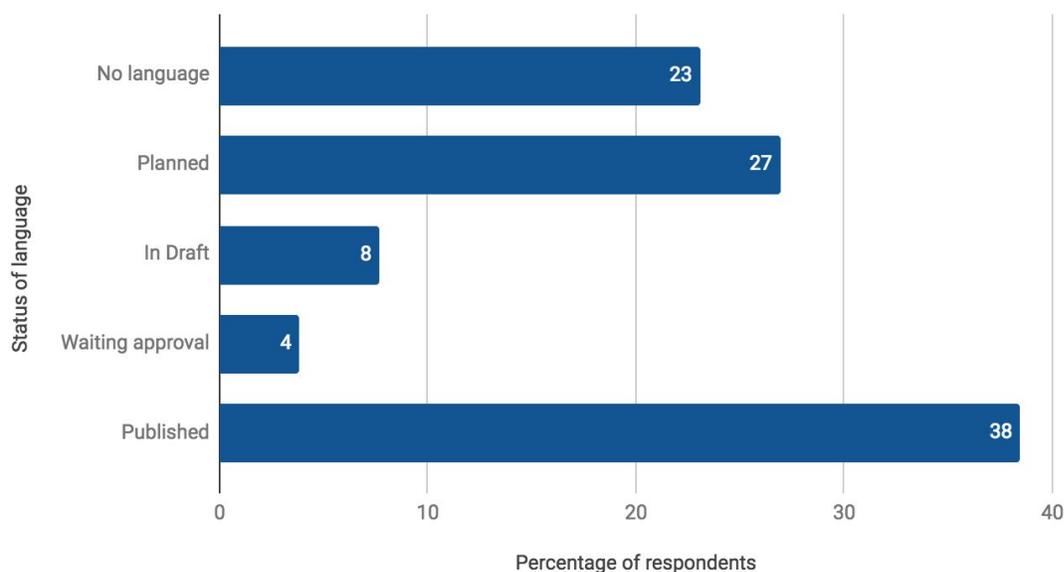
Figure 3: Status of language in a strategic plan or mission/mandate statement expressing commitment to digital preservation by percentage of respondents.

When asked about **areas of activity for digital preservation**, 19 respondents (73%) indicated that it is taking place in archives and special collections units. 11 respondents (42%) said digital preservation is taking place within library systems, information technology or discovery services units and 8 (30%) indicated that it is being undertaken in digital scholarship or digital initiatives units. **23 (88%) respondents have more than one unit or department undertaking digital preservation activities in their organization**.

17 respondents (65%) have an individual or group responsible for coordinating digital preservation activities across the organization. Of these, 7 named a senior individual such as a university librarian, dean, associate dean, or the head of a unit. 3 named a senior leadership team, 3 a working group or committee, and 1 an individual librarian. 3 responded to the question by naming functions of the role, such as coordinating policies and procedures and developing partnerships.

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

Individuals	**Working Groups or Committees**
- Special Collections Librarian - Associate Dean of Libraries, Digital Engagement and Strategy - Associate Dean and the Head of University Archives and Special Collections - Associate Dean, Digital Initiatives	Digital Preservation Working Group (which includes representatives from several units: Archives and Special Collections, Metadata, and Systems. They are responsible for policy development, for designing workflows, and for assigning operational activities like loading material into Archivematica)

● University Archivist, Digital Collections Administrator ● Director of Scholarly Technologies	● The library's senior management and leadership team

13 respondents (50%) have a committee or working group related to digital preservation. Of these, 7 respondents named a formalized group whose work is focused on digital preservation; 2 named an informal group, and 4 others indicated that this role belongs to a group with oversight over digital materials more generally, such as a digital collections or digital scholarship group. The roles for these groups were primarily described as identifying and drafting policies, providing oversight for operations, and ensuring alignment with strategic plans. Individual respondents identified additional activities such as monitoring trends, identifying best practices for digitization, and making recommendations relating to systems or inventorying assets. One respondent commented that "There was a working group struck to consider digital preservation. It has finished its work. It made recommendations, but the inability of the library to provide infrastructure and staffing has created a challenge on follow through."

There is a **large amount of participation in many different organizations, projects and initiatives relating to digital preservation** among CARL members at the regional, national, and international levels. **22 respondents (85%)** indicated participation or membership in regional, national, or international organizations, conferences or projects specific to digital preservation. The following word cloud provides a list of the organizations weighted by the frequency with which they were noted. The highest frequency was 7 times (CARL Portage/FRDR), and the lowest 1 time. It is notable that many of the organizations or projects listed are occurring at the level of regional consortia (engagement with COPPUL or OCUL services, for example) or outside of Canada in the United States and internationally, such as HathiTrust, the Internet Archive, and LOCKSS/CLOCKSS. At the national level, CARL initiatives such as Canadian Web Archiving Coalition and Portage/FRDR (Federated Research Data Repository) were the key items mentioned.

Figure 4: Word cloud of organizations, projects, initiatives or conferences related to digital preservation in which CARL members 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 communication/coordination among stakeholders (14 respondents, 54%)
- Lack of resources to pursue organizational change (14, 54%)
- Lack of high-level organizational commitment or support (11, 42%)
- Lack of resources to participate in collaborative efforts (9, 35%)
- No gaps/challenges (2, 8%)

Additional gaps and challenges noted were: a lack of coordination or prioritization at the organization's highest level (4 respondents), a lack of funding or other resources (3); the problem of developing skills or finding digital preservation expertise (2), the challenges of moving from ad hoc to more formalized processes (2), and the large size and scope of the materials to be preserved (1).

Key takeaways:

- The majority of respondents expressed a commitment to digital preservation in general, which is a necessary first step towards developing policies, resources, and capacity.
- In many cases while there is "in principle" commitment, there are gaps in terms of developing and operationalizing the programs at institutions: the coordination of digital preservation work is beginning, and the presence of working groups or committees looking at the subject is mixed.

- The large majority respondents have multiple units with responsibilities in digital preservation.
- Engagement with external organizations, projects and initiatives relating to digital preservation is strong.
- Key concerns are a lack of communication and coordination among stakeholders and a lack of resources to pursue change at the organizational level.

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 8 organizations (31%) have policies in place or are drafting a policy, while 18 (69%) have not.
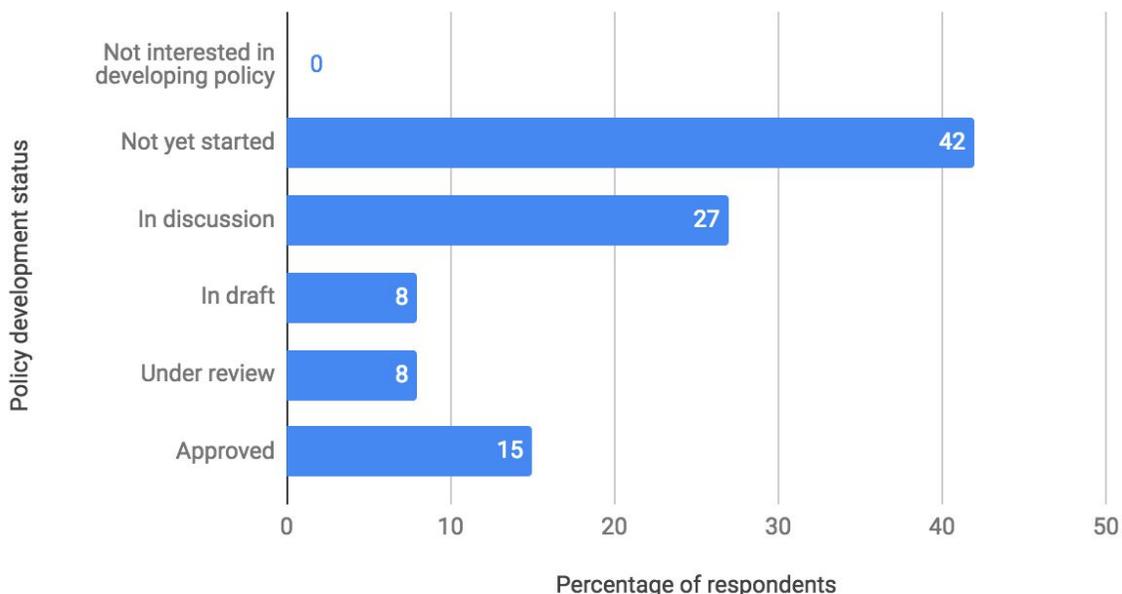

Figure 5: Status of policy development at respondent organizations by percentage of respondents.

Of the 8 respondents who listed a scope for their policies the following topics were noted:
- 4 policies listed treat broad organizational commitments to long-term preservation across the organization's units, including high-level objectives and principles. 2 of these also mentioned ensuring cost-effectiveness/sustainability.
- 2 policies treat applying levels of preservation or prioritizing materials for preservation.
- 1 policy description consisted of a set of specific action plans for different content types owned by specific units, united by a draft framework.
- 1 policy relates specifically to content in a digital repository.

Table 2: Policy scope examples provided by respondents

> "The policy articulates the University Library's commitment to ensuring enduring access to digital objects for our academic community; identify specific objectives and articulate the roles and responsibilities within each unit ensure activities follow accepted and emerging standards."
>
> "This policy applies to all digital preservation activities undertaken within the library system. This policy is intended to outline what can be expected from digital preservation programs and activities and to place reasonable limits on those expectations given expected operating constraints and technological considerations. - This policy may form the basis for partnership agreements if the library provides digital preservation services to non-library partners, now or in the future."
>
> "To identify and prioritize all of the content groups that should be considered for digital preservation. To identify the appropriate level of preservation for each of those groups. To determine and document procedures for each content group. To ensure that selected preservation tools are available, and that the preservation workflows are operationalized."
>
> "The library has a digital preservation policy framework undergoing review. This high-level document, which is based on the seven attributes of a Trusted Digital Repository, highlights the goals, objectives and principles of the library's digital preservation program as well as pointing to documents containing more detailed and frequently-updated procedures, guidelines and the like."
>
> "We have a draft Digital Preservation Framework and eight draft Action Plans, each one documenting our preservation actions for a specific type of content."

The **rated gaps and challenges related to policies** were:
- Lack of time/resources for policy development (20 respondents, 77%)
- Policies are ad-hoc or project specific (13, 50%)
- Lack of knowledge for policy development (10, 38%)
- Policies are not well documented (7, 27%)
- Policies are not well understood or followed (4, 15%)
- Policies are not reviewed (3, 12%)
- Lack of interest in policy development (1, 4%)
- No gaps/challenges (1, 4%)

3 respondents added that a major barrier to policy development is a lack of knowledge among the staff at the institution. One respondent also commented: "Policies need to be at a fairly high level with procedures and guidelines applying to many different types of collections and data types managed via different technologies. This can be challenging to manage."

Procedures and Workflows

Documented procedures and workflows are also not especially well developed among respondent organizations, with most respondents having no procedures, undocumented

procedures, or draft procedures only (together, 22 respondents, 85%). However, the status of procedures in draft (50% of respondents) indicates work is happening in this area.
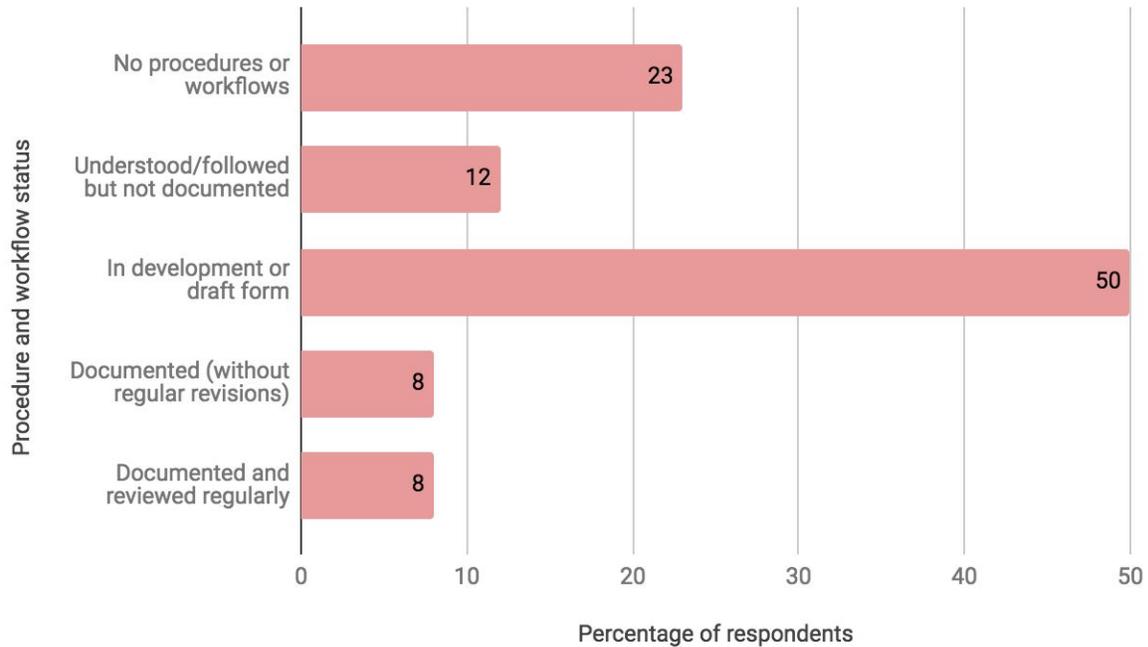

Figure 6: Status of procedures or workflows at respondent organizations by percentage of respondents.

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 (13, 50%)
- Lack of knowledge for procedure documentation (10, 38%)
- Procedures are not well documented (7, 27%)
- Procedures are not well understood or followed (6, 23%)
- Procedures are not reviewed (3, 12%)
- Lack of interest in procedure documentation (2, 8%)
- No gaps/challenges (1, 4%)

Strategies or Plans

15 respondents (57%) have digital preservation plans or strategies in place covering a wide range of content or format types including images, audiovisual materials, institutional repository collections, and digitized content. 3 respondents indicated they have coordinated format guidelines, action plans or registries, such as guidelines or recommendations for researchers on formats accepted. Respondents noted that digital preservation plans or strategies are confined to specific departments, rather than coordinated across the whole organization.

Other Policy-Related Considerations

13 respondents (50%) responded positively to the question, "**Has the institution adopted any particular digital preservation standards, best practices or guidelines?**" The top items listed were PREMIS for digital preservation metadata (5 mentions), OAIS (3), informal use of TRAC/ISO 16363 for self-auditing purposes (3), various digitization standards (2), and file format guidance (2).

9 respondents (35%) have used digital preservation-related self audit or gap analysis tools. The most common tool in use (by 4 respondents) is the TRAC/ISO 16363 standard. Other tools listed one were: DRAMBOA, the Data Seal of Approval, and the Scholars Portal Permafrost digital preservation readiness interview.

When asked about **interest in pursuing formal ISO 16363 Trustworthy Digital Repository** certification, **6 respondents (23%) indicated that they have undertaken or are interested in undertaking formal certification**. The 17 respondents that answered "no" indicated the reasons for not pursuing certification is lack of time/resources (7 mentions), or that lack certification as a priority (5). 3 noted that other self-audit tools fill a similar need, and one respondent indicated that OCUL and Canadiana's TDRs fulfill the major needs of the community. 3 respondents "I don't know" in response to the question.

Key takeaways:
- Policy implementation is limited among respondent organizations. Only a handful of respondents have a digital preservation policy finalized or in draft.
- A higher number of organizations have informal or draft procedures in place, and many make use of strategies or plans. However, these procedures still tend to be ad hoc in nature.
- Standards adoption and other markers of more mature digital preservation programs at the policy level are similarly in progress.
- Difficulties experienced by respondents are determining the scope and role of a policy and lack of time, resources and expertise to develop policies and procedures.

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. **27% of respondents (7) are using tools for digital forensics**. 4 the 7 respondents are using BitCurator and 2 are using Kryoflux controllers alongside a mix of other tools to accomplish this work.

The **rated gaps and challenges related to digital forensics** were**:**
- Lack of staff knowledge/skills (16 respondents, 62%)
- Lack of access to software tools (11, 42%)
- Lack of access to hardware (disk drives, write-blockers, etc.) (11, 42%)
- No gaps/challenges (2, 8%)

Others noted that lack of time and resources (5 respondents), skills and learning (3) and organizational commitment (1) are a barrier in this area.

Preservation Processing Tools

Respondents are split in their use of tools for preservation processing: **half of respondents (13) are using at least one tool for preservation processing and half are not using any tools (13)**. Of the respondents using a preservation processing tool, 11 are using Archivematica as one of these tools. 4 of these respondents are using Archivematica in production and the remaining 7 are in testing. Respondents report a handful of other tools in use: 2 use Islandora's preservation capabilities, and 2 use locally-developed workflows and scripts. In total, 6 respondents (23%) are using any one preservation processing tool in production.

The **rated gaps and challenges rated for preservation processing tools** were:
- Lack of staff knowledge/skills (15 respondents, 58%)
- Lack of software/tool support/resources (12, 46%)
- Lack of money to support tools (9, 34%)
- Lack of access to hardware (6, 23%)
- No gaps/challenges (2, 8%)

In the comments, others noted lack of time (4 respondents), lack of policies (2), or lack of management support (1) to support staff use of tools. Respondents also commented that efforts to implement tools are waiting on policy development and/or a stronger organizational commitment to resource this work. Lack of coordination among institutions putting together workflows for digital preservation was also mentioned as an issue.

6 of the 7 respondent organizations that have digital forensics tools in place are also using Archivematica or other tools for preservation processing; organizational capacity to implement one set of tools clearly leads to the capacity to implement another set.

Key takeaways:
- Few organizations have operationalized digital forensics workflows or are using preservation processing tools beyond testing, meaning that while assets may be stored and accessible to users, they may not be being preserved according to standard practices in the field.

● The challenges to adopting these tools resonate throughout the survey: a lack of resources to develop staff knowledge and fund tools and their uses, and the need for a strong base of policy-driven organizational commitment to drive hands-on work.

7. Discovery and Access

Access is clearly a priority among respondents and they generally expressed confidence in this area. **All 26 respondents noted that they endeavour to provide access to digital objects under their care**. The methods by which they do so are quite varied. While **all respondents use web-based platforms** to disseminate digital materials, other transfer methods such as online shared folders (15 respondents, 58%), dedicated computer terminals (9, 25%), physical carriers (3, 12%), and email (1) were also noted.

DSpace and Dataverse are the most commonly used web-based platforms, followed by Access to Memory (AtoM), Islandora, and ContentDM. Most respondents **use these systems for managing both digital objects and descriptions/metadata**, however 7 of the organizations using AtoM use it to provide access to metadata descriptions only.

Table 3: Access Systems in Use

System	Digital objects only*	Descriptions only	Digital Objects and Descriptions
ArchivesSpace		1	
AtoM		7	10
Blacklight			1
ContentDM			7
Dataverse	4	1	12
DSpace	2		14
Islandora			8
Samvera/Hydra			2
Other	2		5

*May be interpreted to mean uncurated metadata and/or the use of storage repository functions only.

The **rated gaps and challenges related to access** were:
● System/software limitations (14 respondents, 54%)
● Lack of technological infrastructure (11, 42%)
● Lack of policies/procedures (10, 38%)

- Privacy/security issues (7, 27%)
- Lack of storage space (3, 12%)
- No gaps/challenges (3, 12%)

Additional concerns expressed by respondents were related to intellectual property issues and donor-restricted records (4 respondents) and the cost of expanding and improving existing platforms and storage (3). 2 respondents noted there was a problem of accessing data at scale, particularly for researchers interested in text mining and data analytics. One respondent noted, "Big data (eg. petabyte range) does not yet have mature solutions for even the national research computing space. Everything else I feel fairly good about."

Key takeaways:
- Access systems and approaches are relatively mature among CARL members, with a variety of different systems in use.
- Challenges in this area relate to improving systems and infrastructure, and issues of policies and procedures and handling intellectual property and privacy.
- Larger-scale data mining and data analytics are not yet supported by most of the systems in use.

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 this diversity brings its own challenges. While 96% are using local network storage as one option, the **strong use of cloud storage and other replicated network storage options or tape** shows a general commitment to appropriate backups and maintenance.
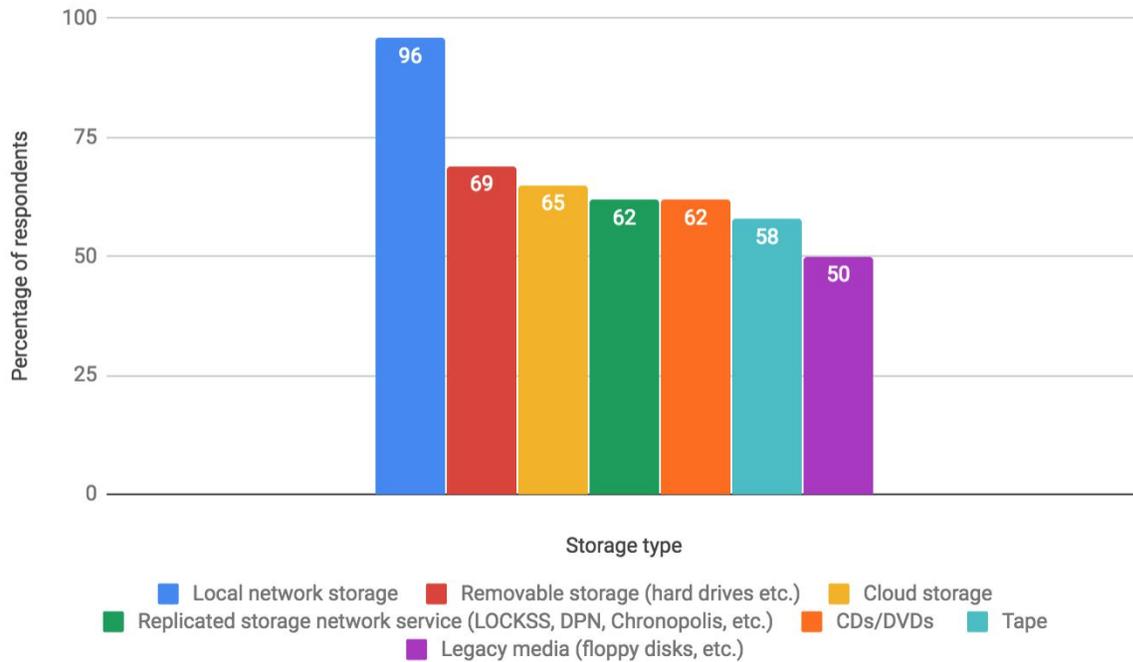
Figure 7: Storage methods selected by percentage of respondents.

Of the 17 respondents (65%) using cloud storage, 8 are using private/community clouds, 4 are using commercial cloud providers, and 4 are using both methods. 1 respondent did not list the type of service used.

All but one of the respondents are using multiple storage methods and media, with the **average number of methods used** being 4.6. Multiple storage places are a good thing when data can be managed well over time, but 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 storage for preservation can be costly at larger scales (especially for audiovisual materials) and good backups and monitoring systems for local storage are required.

Quantities

In responding to the following questions, several respondents noted that it is difficult to determining approximate numbers for quantities and locations for their digital assets, which points to a lack of systems that provide oversight and management of these assets. One commenter said: "I would expect that few libraries that have extensive digital collections can easily determine how much disk storage each type of content consumes. Answering this question would require more sophisticated digital assets management capabilities than most libraries are capable of."

54% of respondents (14) were able to indicate the amount of digitized content their organization stores. The median amount of data was 20 TB, with a low of 2 TB and a high of 175 TB. 6 respondents fell between 0 and 10 TB; 7 between 10 and 70 TB; and 1 between 100 and 200 TB.

38% of respondents (**10) were able to indicate what quantity of born-digital content their organization stores**. The median amount was 10 TB with a low of 1 TB and a high of 330 TB. The latter figure is possibly the result of web archiving activities. 5 respondents fell between 0 and 10 TB; 4 between 15 and 40 TB; and 1 at over 40 TB.

Distributions for Born-Digital Content

65% of respondents (17) were able to estimate the distribution of born-digital data across storage methods. Those who responded indicated that **71% of materials are on networked systems while the rest, approximately 28%, still reside on external media**. Just under 1% are on internal media, such as a personal computer. Both internal and external media may include legacy media like floppy disks and hard drives from private donors that will require forensic imaging for archival purposes. Such activities will require the expansion of digital forensics capabilities as discussed in the "Tools and and applications for preservation" section.

Respondents did not mention these issues when considering challenges relating to storage itself. The **rated gaps and challenges related to storage** were:
- High cost of local storage (12 respondents, 46%)
- Lack of local storage (6, 23%)
- Lack of support for storage (6, 23%)
- Lack of oversight/control over storage (4, 15%)
- Procurement barriers (4, 15%)
- Lack of backups (3, 12%)
- No gaps/challenges (5, 19%)

Several commenters noted the difficulty in organizing the relationships between the library and university information technology providers, and the general problem of underutilization of available storage. But in general, respondents were fairly confident about storage solutions, at least regarding access to storage.

Key takeaways:
- A variety of storage systems are currently in use, many of them including physical media-based storage methods that may be at risk. Resources for transfer of data from these materials, including forensics capabilities, will be required to ensure the safe transfer of these materials to more reliable storage systems.
- Respondents are generally confident in the storage infrastructure available to them.
- Evidence of a transition to preservation-friendly storage is positive.

- There is a clear need for management and oversight of digital assets across storage systems.

9. Staffing

Staffing with responsibilities for digital preservation among respondents is low. There is very little standardization in positions and responsibilities across respondent organizations and a great diversity in the number and scope of roles for staff involved in digital preservation activities. 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

96% of respondents have one or more roles with responsibilities for digital preservation. 27% of respondents (7) have 3 roles with responsibilities for digital preservation, 15% (4) identified 5 roles, and 15% (4) identified more than 5 roles.
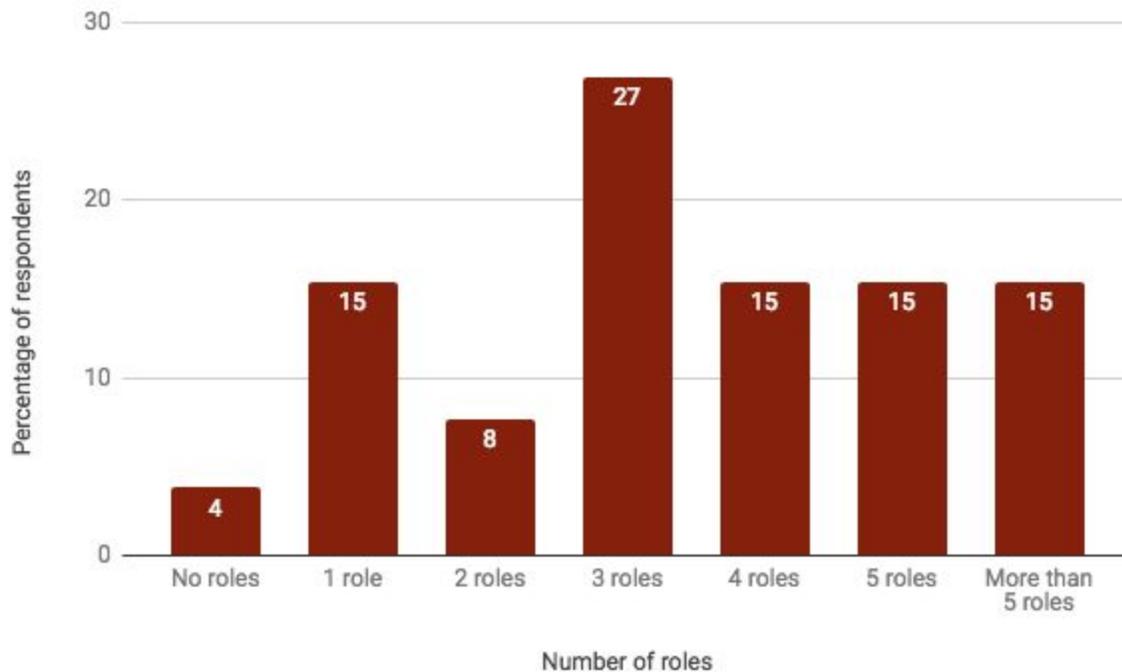

Figure 8: Percentage of respondents by number of roles dedicated to digital preservation.

Job titles for digital preservation-related roles varied widely. Of the 86 job titles entered by respondents, "Archivist" and variations on "Digital Archivist" were listed 10 times. Variations on "Digital Librarian" (Digital Asset Librarian, Digital Initiatives Librarian, Digital Preservation Librarian, etc.) were listed 25 times. "Systems Librarian" was listed 3 times. Titles belonging to

senior administrative staff, such as university librarians or archivists, associate university librarians, and heads of units, were listed 13 times.

FTE Values

Respondent-entered FTE values showed the relatively **low amount of staff time directed toward digital preservation work**:

- 4 respondents (16%) that entered FTE values have at least one 100% FTE role entirely dedicated to digital preservation.
- 5 more respondents (19%) have staff roles that add up to 100% FTE or more.
- Therefore, 65% of respondents have less than the equivalent of one individual working on digital preservation overall.

Figure 9: Average total percent FTE per 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 as well that one respondent was disregarded as they listed job titles but no FTE values.

For respondents with 1-4 roles, the average total FTE across these roles was 70% FTE. When including the 5 respondents with 5 roles, the average FTE rose to 108%, as 2 respondents listed 5 full-time roles.

However, Of the 77 roles with FTE values assigned, 60% (46 roles) were between 0 and 20% FTE. 25% (19 roles) were between 20 to 100% FTE and the remaining 15% (12 roles) were 100% FTE roles.

Overall, total FTE for all roles listed was 2603; the average total per respondent (including respondents with no roles) is 104.12% FTE, or just over one full time position.

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. 1 respondent gave job titles but no FTE values, and 2 respondents gave FTE values for the first role listed but none for subsequent roles.

Expectations for extending staff responsibilities for digital preservation were high. **16 respondents (62%) said they intend to expand staffing**. Of these, 4 intend to do so through new hires, 7 through reassignment, and 5 through both methods.

The **rated gaps and challenges related to staffing** were:
- Lack of funding for new positions (16 respondents, 62%)
- Lack of staff knowledge/skills (14, 54%)
- Lack of resources for training/professional development (5, 19%)
- No gaps/challenges (1, 4%)

Other gaps/challenges identified are difficulties finding qualified staff and difficulties in determining staff needs while a digital preservation program is still in development. One commenter noted: "Il est très long et complexe d'acquérir les connaissances nécessaires pour avoir la base en conservation numérique et aucune formation accessibles facilement ce qui fait qu'un nouvel employé doit souvent être formé à partir de 0 et le champs de la conservation numérique est énorme!" [It is very long and complex to acquire the necessary knowledge to understand the basics in digital preservation and there is no easily accessible training, so that a new employee must often be trained from 0, and the field of digital preservation is enormous.]

Key takeaways:
- Most respondents have low levels of staff devoted to digital preservation. While many organizations have at least one or two individuals working in the area, digital preservation is likely to be only a small portion of their responsibilities. 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.
- The majority of respondents expect to expand staffing.
- The major challenge as stated by respondents is a lack of funding for positions.
- Additional challenges are difficulties in training existing staff or finding new staff with adequate expertise in the field.

10. Funding

81% of respondents (21) are funding digital preservation through general budgets. However, **22 respondents (85%) did not know what percentage of their organization's budget was dedicated to digital preservation**. Of the 4 that did respond, 3 indicated that less than 1% of their budget is dedicated to digital preservation activities and 1 indicated that it is 3% of their budget. In a few cases, preservation funding comes from materials budgets (2 respondents), grants (1), collections budgets (1) and a specific budget line for preservation (1).

While 20 organizations (77%) anticipate that there will be funding increases for digital preservation in the next 1-2 years, they were not able to anticipate what level of increase is expected, and in some cases, this may not indicate significant new funds coming beyond normal, expected increases.

Most respondents indicated that there are additional resources for digital preservation coming from other departments within the university, and also in some cases from external organizations.

Table 4: Sources of external resources for digital preservation activities

Internal departments	External organizations
IT department (18, 69%)	Regional consortium (17, 65%)
Legal department (4, 15%)	National consortium (8, 31%)
Records management and archives (2, 8%)	Professional association (3, 12%)

The **rated gaps and challenges related to funding** were:
- Lack of business plan (14 respondents, 54%)
- Allocation of resources is too low in comparison to needs (13, 50%)
- Funding is not sustained (8, 31%)
- No gaps/challenges (1, 4%)

Some respondents noted that changes in leadership, as well as the "in development" nature of digital preservation functions means that funding remains static. Another said "we are one of those institutions which takes digital preservation seriously and put a lot of resources but still resources are limited in comparison to needs."

Key takeaway:
- General budgets are the main source of funding for digital preservation programs. However, few respondents were able to indicate how much money is devoted to digital preservation work.
- Most respondents expect funding to increase in the next 1-2 years.
- Most respondents receive additional resources internally from IT departments, and externally from regional consortia.

- Funding is inadequate for at least half of respondents. This may be because funding for digital preservation is difficult to estimate, in part because digital preservation activities have not reached maturity among most respondent organizations or fall across a number of units or departments within the organizations, making digital preservation costs difficult to separate from other functions.

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). 10 (38%) felt that they are at the initial stage, and 10 (38%) felt they are one step up up at the repeatable stage. 5 (19%) felt they have reached a defined stage, and 1 respondent (4%) felt they are at the managed stage. No organization indicated that they were at the "no activity" or optimized stages.
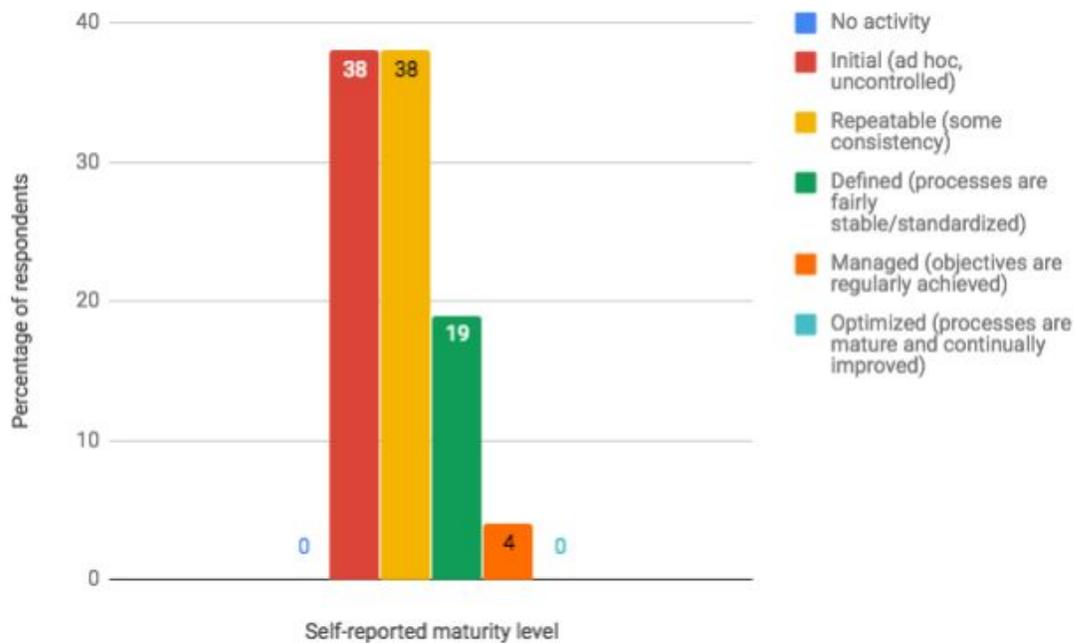

Figure 10: Self-reported maturity scales by percentage of respondents.

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:

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

"Most significantly, we need to create the time to develop policy and procedures. Resources (storage, software) are not significant barriers at this time."   "We need to go slow on consortial efforts to allow for enough time for the collective effort to coalesce."

> "1. Centralized IT service provider does not understand digital preservation requirements and cannot provide solutions needed.
> 2. Organizationally, other parts of organization can override digital preservation decisions and recommendations.
> 3. Funding fluctuates depending on the current management in place."
>
> "Lack of stable funding for digital preservation program.
> Lack of trained and qualified technicians.
> Lack of institutional digital preservation policy."
>
> "It's too early in some ways, to know this. We need to do more work, and do more asking, before we can identify other resource issues."
>
> "Organizationally we have to find a 'home' for this within our team structure."

Key takeaway:
- The majority of respondents (76%) do not yet see a defined or formalized set of digital preservation activities as core to their operations, while a smaller group (23%) 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 this means will vary. Different content types will require different strategies and resources, and preservation approaches are always in flux as technologies change and materials age. An institution may rely upon a consortial preservation program for its journals holdings, and devote more intensive resources to born-digital materials acquired by its archives. Therefore, digital preservation programs are best evaluated based on the sets of functions and activities they are able to support relative to their needs.

The lowest score was 4.5 and the highest was 14; the average score overall was 8.98. The following chart shows the distribution of scores:
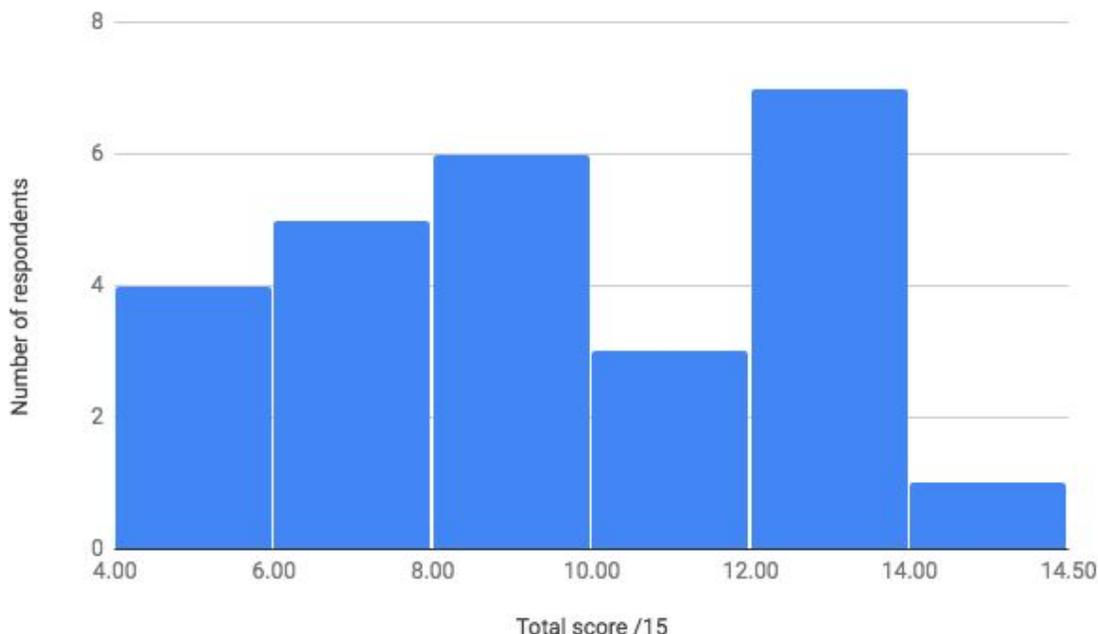
Figure 11: Distribution of scores out of 15 by number of respondents.

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

Score range	% of respondents	% relative capacity
0-4	0%	0-27%
4-6	15%	27-40%
6-8	19%	40-53%
8-10	23%	53-67%
10-12	12%	67-80%
12-14	27%	80-93%
14-15	4%	93-100%

Looking across respondent organizations, some interesting patterns emerge. For example, there is not a large correlation between digital preservation efforts and size of institution (as measured by budget and library staff FTEs).[1] While 2 institutions with relatively high materials expenditures and staffing levels come out on top in terms of digital preservation scores, the other top 10 institutions in terms of scores do not necessarily reflect the highest collections

---

[1] Published CARL statistics for 2015-16 were the basis for these comparisons.

budgets or FTE staff numbers. 3 of the top scoring institutions for digital preservation fall below the median materials expenditure of roughly $10.5 million, and 4 fall below the median staff FTE number of 149. However, on the other end, the bottom 10 scorers generally have expenditures and total FTE levels either around or below the median level. Therefore, while a large institution can help support a strong digital preservation program, it is not a prerequisite.

The top 10 scorers are strong across all functional areas, although 5 of the top 10 respondents have lower scores in the staffing and resources area (which was a weak area overall). The bottom 10 score weakly across the board. And the middle 6 have scattered areas of strength, but also other areas of relative weakness. Because effective digital preservation programs depend on a broad series of interrelated functions, it is important that institutions show strengths across all aspects of the spectrum: policies, staffing, funding and uses of tools and infrastructure. For CARL respondents, while organizational commitment, and access and storage systems are strong areas overall; policy development and preservation technology use scored lower, while staffing and resources was the weakest area for most organizations.
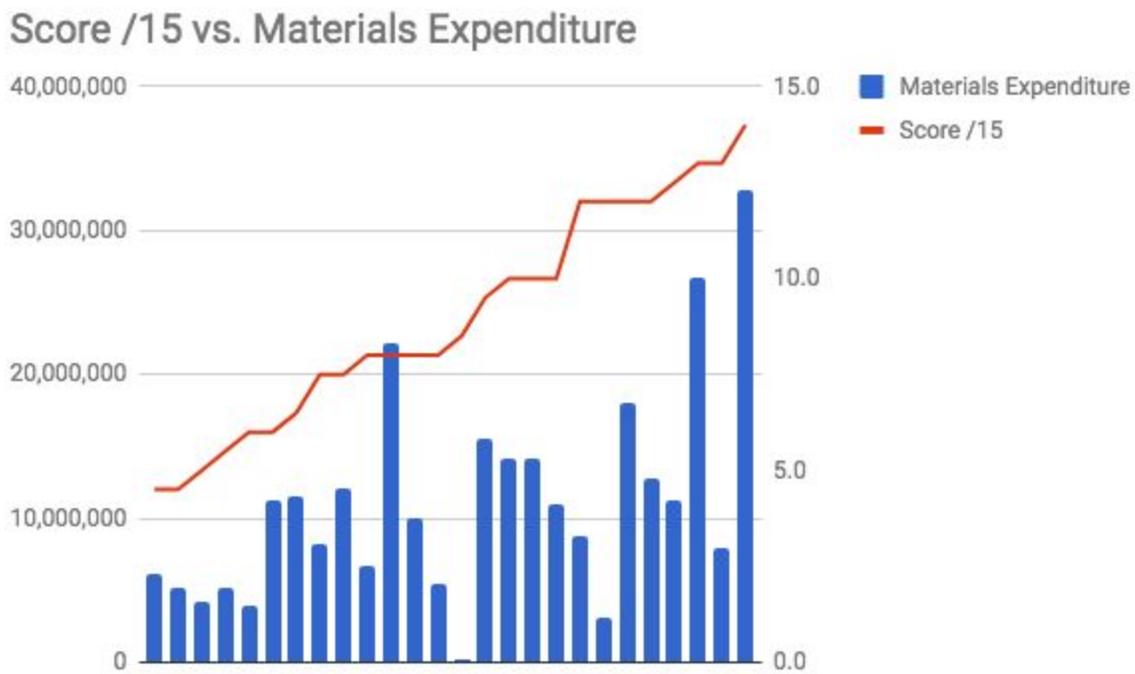

Figure 11: Respondents ranked by score out of a total of fifteen (red line, from lowest to highest) against annual library materials expenditures in millions.

Figure 12 : Respondents ranked by score out of a total of 15 (red line, from lowest to highest) against total library staff FTE.

Overall, it is fair to say that the cohorts of scores roughly map out into thirds:

- Just over one third (9 respondents, 34.5%) have programs that are just starting at about 20-50% capacity. Interestingly, some of these members have strong scores in technology and storage areas but are missing organizational commitment or policy development, and score especially low in available resources and staffing.
- A second set of scores of 50-80% have "in progress" programs (9 respondents, 34.5%) that are in development but missing core pieces, such as policies, uses of tools, and access to sufficient storage.
- A final third of respondent organizations (8, 31%) have generally well developed digital preservation programs with scores of 80-90% capacity overall (when ranked against fellow respondents, rather than against an absolute value for capacity).

As a control, these values map fairly closely to the self-reported maturity areas in the "Concluding comments" section.

13. Conclusion

All respondent organizations are collecting and creating digital materials. And all respondents have a mandate to ensure these materials are accessible into the future. These digital collections are varied and spread across different units and departments within an institution, creating a challenging environment for governance and coordination of digital preservation activities and functions. The majority of respondents expressed a commitment to digital

preservation in principle. However, there are significant gaps between commitment and action for most CARL members. Many organizations do not yet have formal policies, procedures and workflows in place, due to a lack of prioritization or lack of staffing resources. In addition, organizations do not have the expertise needed, both in terms of staff training, as well as finding new staff with the appropriate knowledge and experience. Sufficient resources to do the work, both in terms of staffing and to cover the costs of preservation work, are clearly an issue for many respondents. That said, there is a minority of organizations who have found success in developing digital preservation programs in multiple areas of activity, and therefore have higher relative capacity overall. A more in-depth understanding of their steps towards success would lend greater depth to the survey results.

Overall, there is a sense that most respondents know what they need to do to move digital preservation work forward at their organizations, but issues in one particular area or another create barriers that make it difficult to shift from the present to a desired state. For some, the barrier is due to a lack of clear policies or procedures to guide the work. For others, it sits squarely on a lack of resources to hire staff to do the work. Others noted issues in organizational commitment, which impacts all of these other areas, as the primary challenge. How can organizations address the shared challenges faced by Canadian institutions in the area of digital preservation? One respondent suggests: "We acknowledge that the challenge is bigger than any one institution to address - some level of national coordination or program support ... would be beneficial." The results of this survey, and Phase 2 of the survey that collected information from a broader range of organizations, provide a detailed picture of the capacity for Canadian memory institutions to do digital preservation work at this time. 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 gather information for the survey. There would be no results to report without you! Many thanks go to Lise Brin at CARL for her tireless assistance and support in administering the survey. The authors also thank members of the DPWG for feedback on this report, and support for 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 figure in at least one column - 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.