Results of the CARL DPWG Survey

of

Digital Preservation Capacity and Needs

at Canadian Memory Institutions

Grant Hurley
Digital Preservation Librarian, Scholars Portal & Lead, Survey Subgroup CARL Digital Preservation Working Group

@Risk North 2 | Montréal | November 9, 2018
The Survey

Goals:
• Establish a clearer picture of the state of digital preservation activities in Canada
• Identify gaps in capacity and needs at Canadian memory institutions

Phase 1: members of CARL - 26 respondents

Phase 2: memory institutions (non-CARL members) - 25 respondents

7 academic organizations
5 community/non-profit organizations
13 government-based organizations
Digital preservation is:

“The series of managed activities necessary to ensure continued access to digital materials for as long as necessary.”

[Digital Preservation Coalition Glossary.]
The Results: Outline

- Collections
- Organization and governance
- Policies, procedures, and standards
- Tools for preservation
- Access and discovery
- Storage
- Staffing
- Funding
- Overall strengths and weaknesses
If your organization is collecting/accepting born-digital materials, from where are they sourced?

- Organization itself (records, publications, etc.): 88%
- Private donors: 83%
- Faculty members/researchers: 73%
- Government (publicly accessible materials): 52%
- Purchases from commercial providers: 48%
- Students: 46%
- Harvested web content: 44%
Collections

Content Types

If your organization is collecting born-digital content, what are their content types?

- Still images: 88%
- Audio: 81%
- Moving images: 81%
- Private papers: 79%
- Administrative/operational records: 71%
- Research data: 65%
- Theses: 65%
- Open access publications (books, journals): 63%
- Purchased publications (books, journals): 60%
- Purchased datasets (geospatial, statistical, etc.): 54%
- Databases: 52%
- Government publications and data: 52%
- Web archives or web-harvested materials: 52%
- Artworks: 42%
Collections

Summary

- Common areas of convergence in collections for all respondents: organizational records and private archives; media (photos, audio, video)

- Academic sector places higher emphasis on a wider variety of materials: theses, web archives, purchased/licensed publications, research data

- All respondents are involved in digitization activities

- 3 respondents (5%) indicated that they are not collecting born-digital materials at present.
  - All are government units of different types
  - They commonly noted that they are not yet able to accept born-digital materials due to lack of capacity to preserve them
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?

- 24% no language
- 42% in progress
- 35% published
Organization and governance

● 71% of respondents have digital preservation activities occurring in more than 1 unit
  ○ Most common units are archives and special collections; systems/IT
  ○ CARL members tended to have more units involved (88% CARL vs. 52% non-CARL)

● 76% have an individual or group responsible for coordinating digital preservation activities across the organization
  ○ Majority of individuals listed were teams or senior administrators
  ○ Non-CARL respondents were more likely to have an individual librarian or archivist in a non-management role responsible (12% CARL vs. 36% non-CARL)

● 47% have a committee or working group for digital preservation
71% of respondents are engaged in an external organization, project or initiative related to digital preservation

- 85% of CARL respondents vs. 56% of non-CARL engaged
- Community/non-profit engagement 20%; Government respondent engagement 53%

Phase 1 & 2 word clouds for organizations named
Organization and governance
Gaps and challenges

- Lack of resources to pursue organizational change - 65% selected
- Lack of resources to participate in collaborative efforts - 49%
- Lack of communication/coordination among stakeholders - 43%
- Lack of high-level organizational commitment or support - 43%
Organization and governance

Summary

- There is a broad commitment to digital preservation among respondents.

- Most respondents have an individual or group responsible for coordinating and authorizing work in digital preservation, though these activities may be occurring across many different units of an organization and therefore could be competing for resources.

- CARL respondents are engaged in the larger arena of initiatives, projects, and conferences that are currently active in this space; respondents from other sectors are less likely to be connected to these networks.
Policies, procedures, and standards

Policy

What is the status of your organization’s digital preservation policy?

- 33% not started
- 49% in progress
- 18% approved
- 42% of CARL respondents vs. 24% of non-CARL respondents said “Not yet started”
Policies, procedures, and standards

Procedures

What is the status of digital preservation procedures or workflows?

- 18% not started
- 59% in progress
- 24% documented

- 8% of CARL respondents vs. 28% of non-CARL selected “Documented and reviewed regularly”
Policies, procedures, and standards

- 63% have **digital preservation-related plans or strategies** in place
  - e.g. guidelines for format migration, retention, digitization

- 53% have adopted common **standards, best practices or guidelines in digital preservation**
  - e.g. OAIS, METS/PREMIS, etc.

- 27% have used a **self-audit or gap analysis tool**

- 18% are interested in achieving **Trustworthy Digital Repository certification**
  - 23% of CARL members vs. 12% non-CARL; of these two were government-based archives
Policies, procedures, and standards
Gaps and challenges

Policy
- Lack of time/resources for policy development - 80% selected
- Policies are ad-hoc or project specific - 41%
- Lack of knowledge for policy development - 29%

Procedures
- Lack of time/resources for procedure documentation - 75% selected
- Procedures are ad-hoc or project specific - 47%
- Lack of knowledge for procedure documentation - 31%
Policies, procedures, and standards

Summary

- Formalized and documented policies in progress

- Slightly stronger status of documented procedures and strategies points to work motivated by practical concerns

- Adoption of standards is mixed

- Main issue appears to be resources/time and expertise to devote to policy/procedure development

- Little appetite for formal certification
Tools for preservation

- 27% using tools for digital forensics
  - Of non-CARL respondents, the majority of these users were government-based

- 41% using tools for digital preservation processing
  - Of the 21 respondents using tools, 81% are using Archivematica as one of those tools
  - Of Archivematica users, 59% are using it in a testing capacity and 41% in production
  - Other tools: Preservica, locally developed/Apache Taverna-based workflows, Islandora
Tools for preservation
Gaps and challenges

Digital forensics
- Lack of staff knowledge/skills - 57% selected
- Lack of access to software - 43%
- Lack of access to hardware - 39%

Digital forensics
- Lack of money to support tools - 55% selected
- Lack of staff knowledge/skills - 55%
- Lack of software/tool support - 53%
Tools for preservation

Summary

- Low adoption of tools overall
  - A minority of organizations (fairly evenly across sectors) are using forensics and processing tools
Access and discovery

- 100% of CARL respondents use a web-based platform to give access to digital objects vs. 68% of non-CARL members
  - Other approaches include shared folders (49%), dedicated computer terminals (40%)
  - Common platforms for CARL members: AtoM, Dataverse, DSpace, Islandora
  - Common platforms for non-CARL members: AtoM + mix of various open and proprietary products
Access and discovery
Gaps and challenges

- Lack of infrastructure – 51% selected
- System/software limitations - 51%
- Lack of policies/procedures - 41%
Access and discovery

Summary

- Access is generally more mature area, though more consolidation/integration of tools may be needed
- Respondents noted in comments that privacy/intellectual property issues, as well as scalability for existing tools a barrier
Storage

Storage systems/media use

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

- CARL members are higher users of cloud storage (65% vs. 28% non-CARL)
Storage

Storage quantities

- **Median for digitized content**: 8.5 TB
  - Low: 0.5 TB
  - High: 175 TB
  - Average: 28.54 TB

- **Median for born-digital content**: 3.95 TB
  - Low: 0.05 TB
  - High: 330 TB
  - Average: 27.66 TB

37% of respondents responded “unknown” across the board
Storage

Storage distributions

Of the 61% of respondents who responded:

- 61% of assets on networked systems
- 2% of assets on internal media (e.g. donor’s personal computer)
- 37% of assets on external media (magnetic, optical, hard drives, flash)
- 39% of respondents overall indicated “unknown” or response did not add to 100%
- Community/non-profit respondents largely have assets on external media (93%)
Storage
Gaps and challenges

- High cost of local storage - 51% selected
- Lack of local storage - 39%
- Lack of support for storage - 25%
- Lack of oversight/control over storage - 25%
Storage

Summary

● More confidence in storage in comparison to other areas
  ○ But transition to cloud or other replicated storage networks has been slow, especially outside of larger academic libraries

● Relatively large proportion of assets are still being kept on fragile external media
  ○ This is especially the case for smaller organizations, some of whom depend on hard drives or other removable media to store assets
Average number of roles per respondent: 3
Average portion of digital preservation responsibilities per role: 37%
Staffing
FTEs for digital preservation

CARL Members:

- 84% do not have at least one full-time role dedicated to digital preservation
- 65% do not have staff roles with digital preservation responsibilities that add up to 100% FTE
- 2 respondents have 5 or more full time roles

non-CARL Members:

- 83% do not have at least one full-time role dedicated to digital preservation
- 61% do not have roles with digital preservation responsibilities that add up to 100% FTE
- 2 respondents have 5 or more full time roles
Staffing
FTEs for DP

Average FTE for digital preservation among CARL members: 1.04

Average FTE for digital preservation non-CARL members: 1.19

Total Average FTE for digital preservation: 1.11
Staffing

- 47% said that they expected to expand staffing
  - Of these, 75% said they would do so via reassignment
  - 63% also said they would hire new staff
Staffing
Gaps and challenges

- Lack of funding for new positions - 75% selected
- Lack of staff knowledge/skills - 45%
- Lack of resources for training/professional development - 33%
Staffing

Summary

- Low staffing for digital preservation overall, and most staff are focusing on digital preservation as a small portion of their responsibilities
  - Of the 145 positions listed with FTE values, 54% of these had less than 20% FTE for digital preservation

- Intentions to reassign current staff will require training resources; hiring new staff will require attention to what digital competencies are available in educational programs

- Minority of institutions (8%) with stronger staffing component at 5 100% FTE
Funding

- Majority are funding digital preservation through general budget lines
  - Second most common funding source for non-CARL respondents is grants/awards

- 78% **could not indicate** what percentage of their organization’s budget was dedicated to digital preservation
  - Of the 11 who responded, 45% said “less than 1%”

- 65% **expect funding increases** in the next 1-2 years
  - Few could indicate what the level of increase would be
Funding
Gaps and challenges

● Allocation of resources is too low in comparison to needs - 53% selected
● Funding is not sustained - 51%
● Lack of business plan - 41%
Funding

Summary

- Funding inadequate for half of respondents
  - Dependence on short-term money for long-term activities a concern

- Respondents unaware of how funding is allocated
  - Costs difficult to estimate, especially across units
  - Lack of maturity overall contributes to informal/unsustained funding or lack of ability to advocate for funding
Strengths and weaknesses

● Used a scoring rubric for selected quantitative values

● Not an assessment, but a method of seeing strengths/weaknesses at a higher level and understand distribution of capacity

● The concept of “capacity” in this context is only in relation to the data itself; not an external standard
Strengths and weaknesses

Organization and governance: 2.20
Policies, procedures, standards: 1.85
Tools for access and preservation: 1.52
Storage: 2.04
Staffing and funding: 1.16

Total average/15: 8.78
Strengths and weaknesses

<table>
<thead>
<tr>
<th>Score range</th>
<th>% of respondents</th>
<th>% relative capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3</td>
<td>0</td>
<td>0-20%</td>
</tr>
<tr>
<td>3-5</td>
<td>8%</td>
<td>20-33%</td>
</tr>
<tr>
<td>5-7</td>
<td>14%</td>
<td>33-46%</td>
</tr>
<tr>
<td>7-9</td>
<td>29%</td>
<td>46-60%</td>
</tr>
<tr>
<td>9-11</td>
<td>25%</td>
<td>60-73%</td>
</tr>
<tr>
<td>11-13</td>
<td>16%</td>
<td>73-86%</td>
</tr>
<tr>
<td>13-15</td>
<td>8%</td>
<td>86-100%</td>
</tr>
</tbody>
</table>
Thank you for listening to the numbers!

Questions? Get in touch! grant@scholarsportal.info