An Annual Survey Towards a Collaborative Digital Preservation Strategy for Canada

allana.mayer@mail.mcgill.ca → @allanaaaaaaaa
Existing datasets

CHIN (2011) Digital Preservation Survey

NDSA (2013) Web Archiving in the United States

Methodology

- Anonymous, fully optional, 50-question survey in Google Forms
  
  **Limitations:** long, detailed, technical, cloud-hosted

- Promoted yearly during October
- Survey promoted on ARCAN-L, CARLIS-L, AAO, AABC, ASA, and snowballed on social media

  **Limitations:** informal, unrigorous (self-selection / tech divide), impersonal, uneven distribution due to closed lists
Methodology

- Collected institutional names and staff titles for de-duplication
- Provided open comments section on each page for elaboration
- Changes in 2014: added staff title options; added question about fixity; added “cloud software” to access options
Responses

2013:
27 responses
- 4 international deleted
- 3 duplicates deleted
= 20 used

2014:
27 responses
- 3 international deleted
- 0 duplicates deleted
= 24 used
Demographics

- Municipal / Government: 35%
- Special / Non-profit: 27%
- Academic: 20%
- Private / Corporate: 6%
- Religious: 6%
- Museum: 4%
- Public Library: 2%
Do we know what we have?

One institution, three staff of comparable authority, three responses:

<table>
<thead>
<tr>
<th>Total holdings</th>
<th>Total digital holdings</th>
<th>Processed digital holdings</th>
</tr>
</thead>
<tbody>
<tr>
<td>over 1 million items</td>
<td>Fairly certain (70-99%)</td>
<td>2.90 TB</td>
</tr>
<tr>
<td></td>
<td>Fairly certain (70-99%)</td>
<td>1.74 tb</td>
</tr>
<tr>
<td></td>
<td>Guessing (0-10%)</td>
<td></td>
</tr>
<tr>
<td>1.2 million images (700,000 born digital holdings); 2,245 linear metres textual records; 80,000 audio-visual documents (films and sound recordings); 200,000 electronic files (textual records)</td>
<td>Reasonably certain (40-69%)</td>
<td>100 terabytes</td>
</tr>
<tr>
<td></td>
<td>Reasonably certain (40-69%)</td>
<td>12 terabytes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completely certain (100%)</td>
</tr>
<tr>
<td>&gt;100,000 files</td>
<td>Completely certain (100%)</td>
<td>2 TB</td>
</tr>
<tr>
<td></td>
<td>Fairly certain (70-99%)</td>
<td>1 TB</td>
</tr>
<tr>
<td></td>
<td>Fairly certain (70-99%)</td>
<td></td>
</tr>
</tbody>
</table>
Do we know what we have?

**Total Holdings**

- **2013**
- **2014**

**Born-Digital Holdings**

- **2013**
- **2014**
Do we know what we have?
How much are we holding?
“We are **not able to acquire any born-digital records**, due to technical and financial considerations. In the past 3 to 5 years there has been a sharp rise in the volume of born-digital records created, received and maintained by our [records-creators]: minutes, spreadsheets, powerpoints, email, blogs, websites, videos, photos, audio, Twitter, Instagram ... **So, we have a big problem!**

“Meanwhile, I know that [our records-creators] are creating born-digital records on a daily basis. Right now, **we advise them to print off their minutes, summary financials, photos, emails, etc.** as we do not have the financial or technical capacity to acquire, preserve, or make available born-digital records.”
“Our born-digital holdings are quite small. We hold one multimedia fonds, and dozens of CD and DVDs. **We are not actively pursuing the acquisition of born-digital records at this time.**

“We are building an open source digital repository (Archivematica and AtoM) and **expect to acquire born-digital records more regularly in the next year or two.”**
What are we holding?
CHIN DPS (2011)
“We just started accepting born digital items and mostly we receive photographs in JPGs or TIFFs. I wouldn't know what formats to convert these to and I'm fairly certain that these formats are acceptable for the time being.”
How often is it asked for?

How often do you field requests for born-digital holdings?

- Multiple times a day
- Once a day
- Multiple times a week
- Once a week
- Multiple times a month
- Once a month
- Multiple times a year
- Once a year
- Never

2013 vs. 2014
More about terminals …

Which operating system(s) and version(s) do you offer access to? Which software environment(s) do you offer access to?

Have you ever installed an operating system, software, or software version as a result of a researcher's request? If so, which one(s) did you install?

How often do you receive requests for installation of an operating system, software, or version?
How often are you unable to access a born-digital holding because of lack of technology?

Which types of software license do you hold for your computer terminals? Do you provide internet access on your computer terminals for researcher use? Do you provide access to cloud-computing technologies?
“We have many in-house computers that researchers can use for any purpose, including accessing a website where we host most of our photograph collection, which includes about 80 born-digital images. We have none specifically designated for providing access to born-digital holdings.”
“Our public access work-stations are stand-alone - not networked, for protection, but that results in a bit of a pain for loading anything onto them....

“Without additional resources I am very reluctant to take on digital preservation on top of everything else... but meantime the born-digital material is coming in, and I'm feeling irresponsible for feet-dragging. I know what needs to be done, but don't know what should be dropped to spend time developing a proper digital preservation strategy.

“Big question for a 1-person municipal archives - which is our higher need: a reference archivist or a digital archivist?

“I know which is cheaper and would have more immediate pay-back- the reference archivist; but we really need a digital archivist too - and those two skill-sets are rarely combined.”
How are we preserving?
CHIN DPS (2011)
PARS/ALA/ALCTS (2013)
“Access, preservation, migration of born digital files is a concern, however, I can barely keep up with daily demands to take action at this time.

“I have inherited Minutes in digital form: just organizing them by file name format and making sure the name matches with the dates of the files takes time.

“Our computers are on an external server so backup and loss of data is not a concern. I worry about the migration of files when we update software.”
Lots of work to do

Would you say you have an established policy for preserving born-digital holdings?

2013:
3 “in process”

2014:
1 “in process”
1 “procedures, no policies”
CHIN DPS (2011)
“[Conversion, migration, and copying] are done randomly according to requests and demands for various types of records. We, as of yet, do not have a policy applied to all.”
The takeaways ...
“Born-digital records are not consistently tracked when accessions are processed and described, making it difficult to answer these questions with precision.

“Our university has established several positions within the last ten years in response ... to the growing amount of born-digital information available from individuals and organizations ... and the research material of faculty that must be preserved and shared due to the requirements of grant programs.”
“I guess I have some work to do.

“In our small archive, myself being largely untrained, digital records management is something I know I need to address but I try really really hard to avoid even thinking about because it **gives me nightmares.**”
“We realize that it is important to have a born-digital preservation policy, but with so few born-digital holdings, we haven't seriously looked into it yet.

“We would be interested in acquiring more born-digital material, but so far no one has offered it to us!”
“The one full-time staff [dedicated to born-digital work] is a Project Lead position for a three-year pilot project to develop a digital repository. Our hope is to continue the funding and convert the project into an on-going program, but no guarantees at this point (we're still in year 1 of the project).”
Next steps ...

● Improve the survey
  ○ direct promotion / higher response rate / repeat responses

● Reports and recommendations
  ○ Skills-sharing or a decentralized service coalition
    ■ policy and procedures? LOCKSS? emulation?
  ○ Collaborative advocacy
    ■ legislation? funding? oversight?
“This whole topic stresses me out!!! :)”