Building bridges for digital preservation in Canada – From theory to practice

Grant Hurley, Digital Preservation Librarian, Scholars Portal

Guest Lecture for GLIS 642 | September 22, 2020

Who is talking?

- My name is Grant Hurley
- I am speaking from Toronto, the traditional territory of many nations including the Mississaugas of the Credit, the Anishnabeg, the Chippewa, the Haudenosaunee and the Wendat peoples. Toronto is now home to many diverse First Nations, Inuit and Métis peoples
- I work as the Digital Preservation Librarian for Scholars Portal, the information technology service provider for the Ontario Council of University Libraries (OCUL)

Previous slide photo: Cropped from "Drawing, Bloor Street Viaduct, Don Section." City of Toronto Archives, fonds 200, series 372, subseries 10, item 189.

What does my job involve?

- “Service provider” is a bit different than “Archivist” - I provide digital infrastructure, tools and support for libraries and archives
- I am responsible for preservation of licensed scholarly materials (journals, books) collectively purchased by OCUL members and other shared collections/platforms hosted by Scholars Portal
- I run the Permafrost hosted preservation service for member-specific collections
- I support the Ontario Library Research Cloud for preservation storage
- I offer training opportunities, advice and assistance to members to support their digital archives and preservation needs
- I participate in networks, committees and research devoted to collaborative digital preservation work

Digital preservation is?

Uh oh

“The series of managed activities necessary to ensure continued access to digital materials for as long as necessary” (Digital Preservation Coalition, 2016)

- If you remove the word “digital,” this definition could easily apply to all analogue materials.
- What does this definition mean in practice?

Some broad digital preservation thoughts

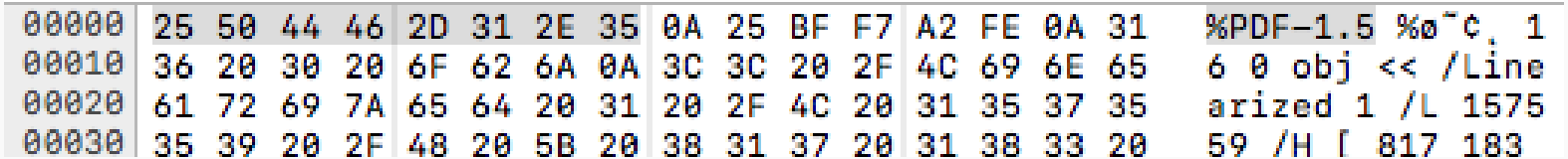
- The broad goals of digital preservation do not differ from the broad goals of analogue preservation: keeping recorded memory alive into the future

- Digital preservation involves the specific practices unique to keeping digital materials authentic, available and reliable over time

- What makes digital materials different?
  - Ease of exact replicability/duplication
  - Many layers of mediation between the object and your ability to access it
  - Complex relationships between materials may make deciding what to preserve complex - such as for websites, video games, etc.

Digital preservation is math

Where the basic functions of physical preservation are chemical (keeping paper, photographs, etc. in the correct storage conditions), the basic functions of digital preservation are mathematical, starting from the binary system of 0s and 1s that forms the basis of all files.



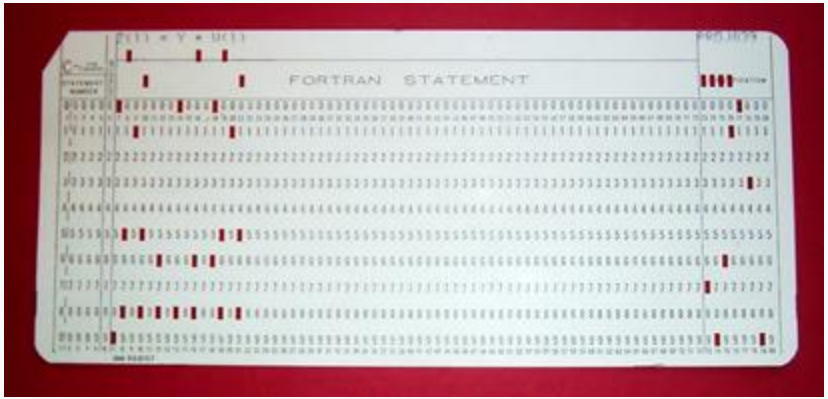
00000	25	50	44	46	2D	31	2E	35	0A	25	BF	F7	A2	FE	0A	31
00010	36	20	30	20	6F	62	6A	0A	3C	3C	20	2F	4C	69	6E	65
00020	61	72	69	7A	65	64	20	31	20	2F	4C	20	31	35	37	35
00030	35	39	20	2F	48	20	5B	20	38	31	37	20	31	38	33	20

%PDF-1.5 %ø¯¢ 1															
6 0 obj << /Line															
arized 1 /L 1575															
59 /H [817 183															

PDF file opened in a in hex editor; the first value (“25”) in binary is 100101; which converts to the character ‘%’ in ASCII text encoding.

Digital preservation is math

The job of the digital preservationist is to know how to interpret this code by knowing its encoding and/or file format and the software that can convert the code back into something useful

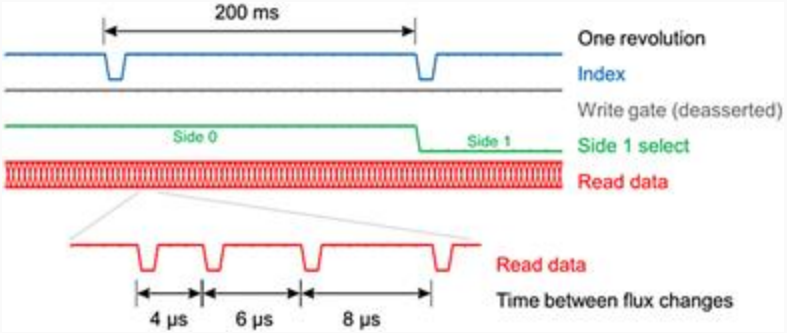


All digital objects are physical

They must be stored somewhere:

- Media (CDs, floppy disks, USBs)
- A hard drive, local server/network share
- The “Cloud” - i.e. someone else’s server

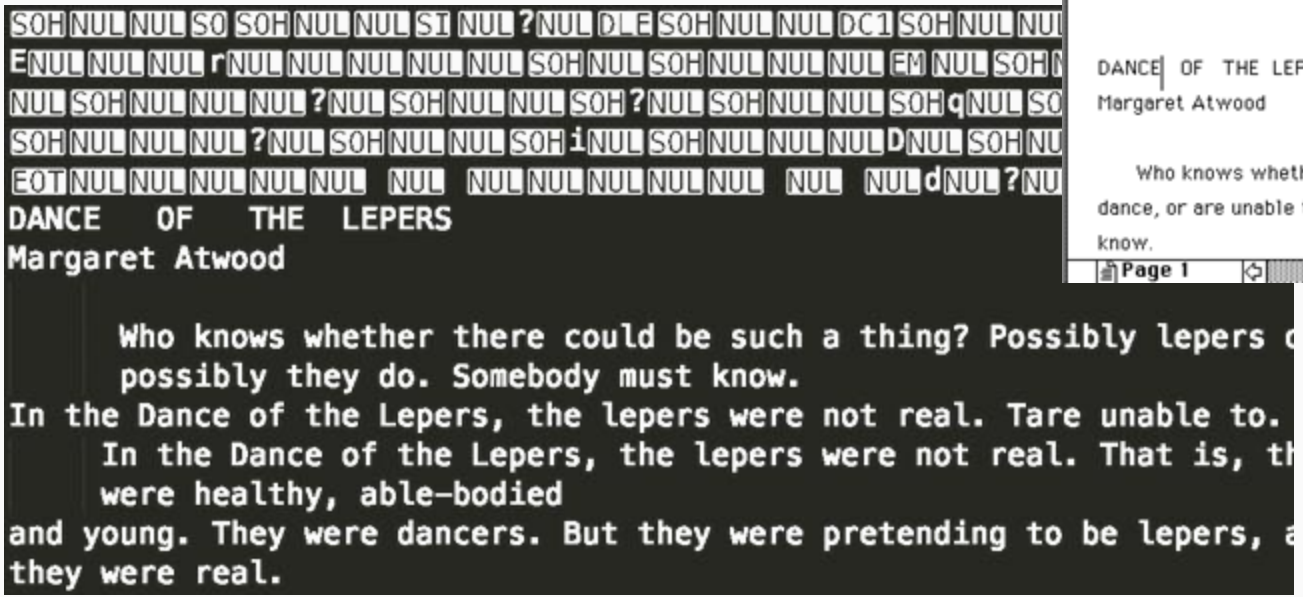
All storage methods physically encode data somehow (punches, divots, magnetic or electric charges) and this takes up physical space

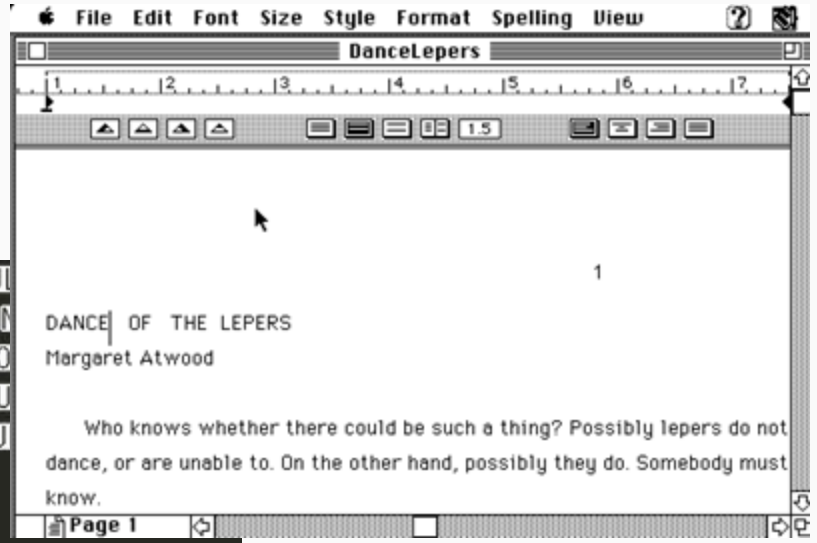


Magnetic modulations on a floppy disk

Digital preservation is not a binary state, even if digital materials are made up of binary code

- A file is not “preserved” - it is intentionally maintained
- Preservers are focused on ensuring that materials remain accessible despite changes in file formats, operating systems, and storage media
- Some materials have better preservation prospects than others
  - An image in TIF or JPEG format or a PDF file vs. a unique file format created by a proprietary software that is no longer available
- … and therefore some materials may require more complex interventions than others

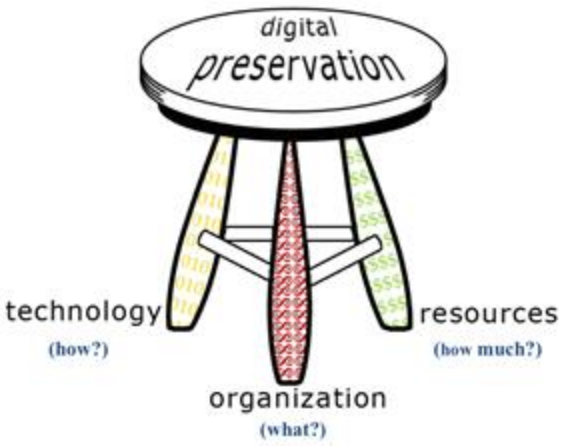




Margaret Atwood manuscript “Dance of the Lepers” in my computer’s text editor vs. emulated in a version of MacWrite II. From the Thomas Fisher Library’s Margaret Atwood papers, collection 547, box 22, disk 1. 1990.

Digital preservation requires resources

- Not only technical infrastructure (tools, hardware, storage), but:
- People with skills to do the work
- An organization committed to the long-term preservation of the materials under its care
- Policies and procedures to guide the work
- Money to pay for all this stuff



The digital preservation three-legged stool by Anne R. Kenney and Nancy Y. McGovern

Key activities

- Acquiring physical custody of the materials to be preserved from some source - and doing so without harming them in the process
- Deciding what materials to retain permanently, and how they should be best accessed, to support the needs of your user community
- Gathering and structuring information about the materials to help inform their preservation, and any future preservation decisions
    - Descriptive metadata; rights metadata; technical metadata (including file format identification, characterization and validation metadata), checksums and more

Key activities con’t

- Processing materials for preservation into AIPs (Archival Information Packages) for long-term storage or DIPs (Dissemination Information Packages) for access
- Storing AIPs safely - multiple copies, multiple locations as resources allow
- Maintaining and monitoring AIPs
  - File formats, fixity status - not just backup-and-forget

“Community archives, Community clouds” paper

- Originally written for digital preservation class at UBC in 2015
- Received the ACA’s Gordon Dodds Prize in 2015 and was published in *Archivaria 81* (2016)
- Helped me get my current job, where I do the things I wrote about in the paper

Still mostly agree with this

“Decades of research have resulted in consensus on the key principles for digital preservation through theory and case studies treating the concept of the digital record, establishing authenticity, defining the needs of different records-creating organizations and individuals, and determining appraisal methods. The practice of digital preservation has also been extensively treated as agreement grows about the use and application of preservation standards, metadata schemas, and data models” (p. 130)

These questions are still generally relevant

“The greatest unanswered questions are the interlocked issues of ‘who’ and ‘where’:

- What scales of layered systems, software, hardware, and people should be maintained for digital preservation, and by which archival organizations?

- How much standardization is necessary so that these systems can speak to each other?” (p. 130)

Developing factors for doing digital preservation

- Growth of shared storage, hosting infrastructures, collaborative networks
  - OLRC
  - WestVault
  - Commercial options
- Maturation of open source tools for preservation processing and access
  - BitCurator (acquisition)
  - Brunnhilde (appraisal, accessioning)
  - Archivematica (preservation processing)
  - AtoM (access)

Current state of activity in Canada

Final Report of the Survey on Digital Preservation Capacity and Needs at Canadian Memory Institutions, 2017-18

for the Canadian Association of Research Libraries' Digital Preservation Working Group

- Publication of final report (English & French) & accompanying dataset

CARL DPWG Report

- Areas of investigation:
  - Collections
  - Organizational commitment
  - Policies/procedures
  - Tools/applications
  - Access
  - Storage
  - Funding

- 52 respondents
  - 34 academic libraries/archives
  - 13 government-based organizations
  - 5 community-based or non-profit organizations

3 challenges + 2 good things

Everyone holds digital collections, but those collections are at risk

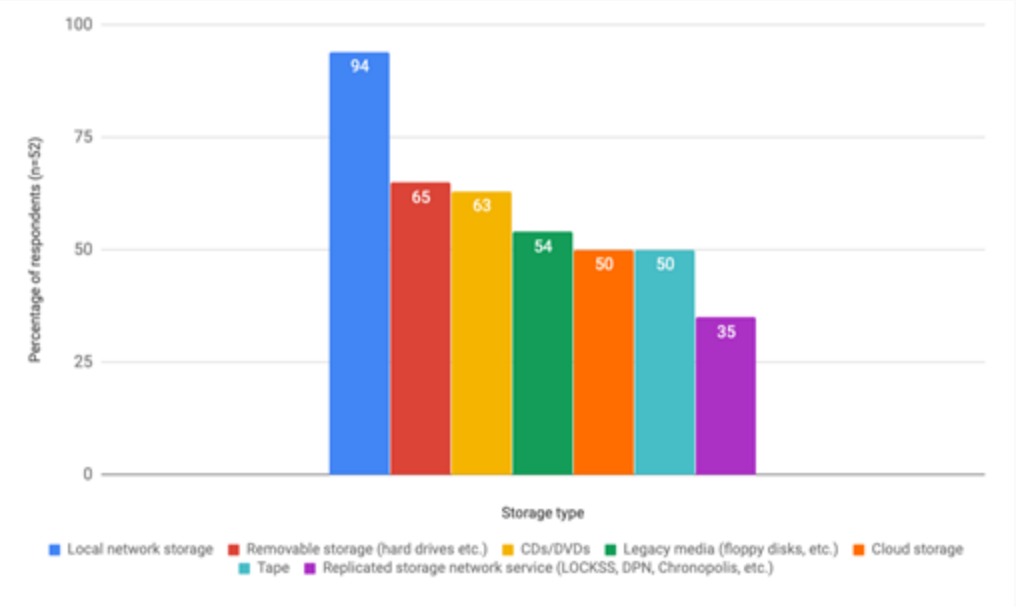
Collections at risk

- All respondents have digitized materials and 94% are collecting born-digital materials

- The median amount of data in storage per organization was **20 TB**

Collections at risk

- Most collections are being stored on local servers or external media
- Fewer institutions are using preservation-friendly storage methods (cloud, tape, replicated networks)



Collections at risk

- 21% of respondents using any one tool for preservation processing in production
- But 96% provide access to digital materials under their care
- Problem: if you lose it, you can’t provide access to it

Very few people are doing this work in Canada

Low staffing

On average, staff with responsibilities for digital preservation represent 0.77% of all staff FTEs at respondent organizations

For every 100 staff members at an organization, less than 1 of them is supporting digital preservation work

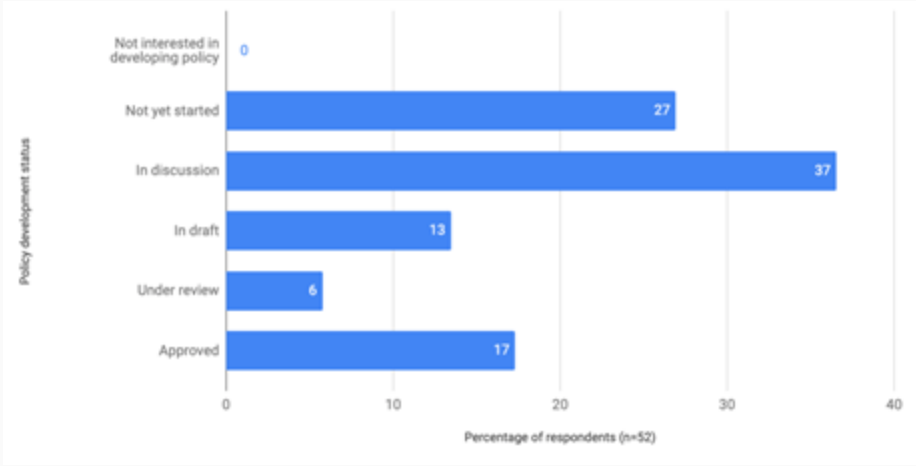
Low staffing

- 38% of respondents have the equivalent of one full-time individual working on digital preservation across **all staff** listed with responsibilities in this area
- 18% have at least 1 full-time individual supporting digital preservation work
- But 48% said they expected to increase staffing through new hires or reassignment

Policies and procedures are lacking

Policies and procedures are lacking

- 17% have published digital preservation policies
- 23% have procedures that are documented
- More policy/procedures in draft form



Top “gaps and challenges” selected

1. Policy: Lack of time/resources for policy development 79%)
2. Staffing: Lack of funding for new positions (77%)
3. Procedures and workflows: Lack of time/resources for procedure documentation (75%)
4. Organization: Lack of resources to pursue organizational change (63%)
5. Tied:
    a. Forensics: Lack of staff knowledge/skills (54%)
    b. Preservation processing tools: Lack of money to support tools (54%)

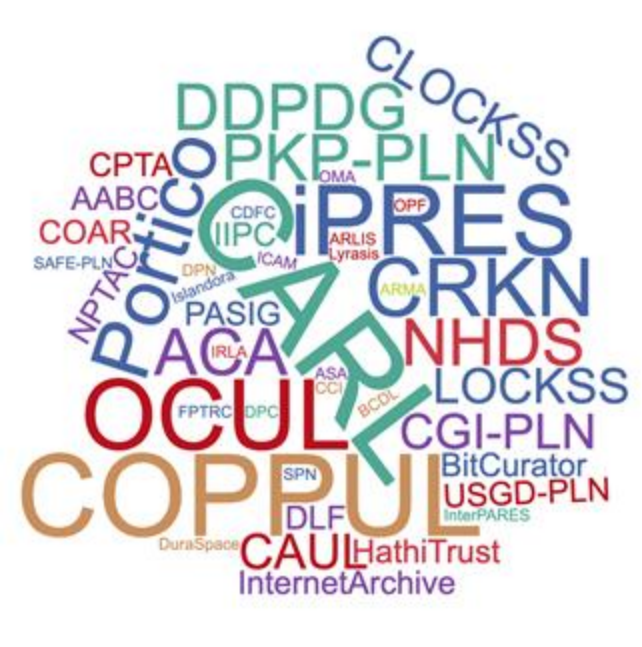
Some good news

Collaboration

There are many organizations and services ready to support this work

- 75% indicated participation in regional, national, or international organizations, conferences or projects related to digital preservation

- There are lots of collaborative things happening:
  - 49 distinct entities were mentioned by respondents
  - 44% of these were mentioned just once



Examples

International

- Digital Preservation Coalition, Open Preservation Foundation
- iPRES, PASIG conferences

National (Canada)

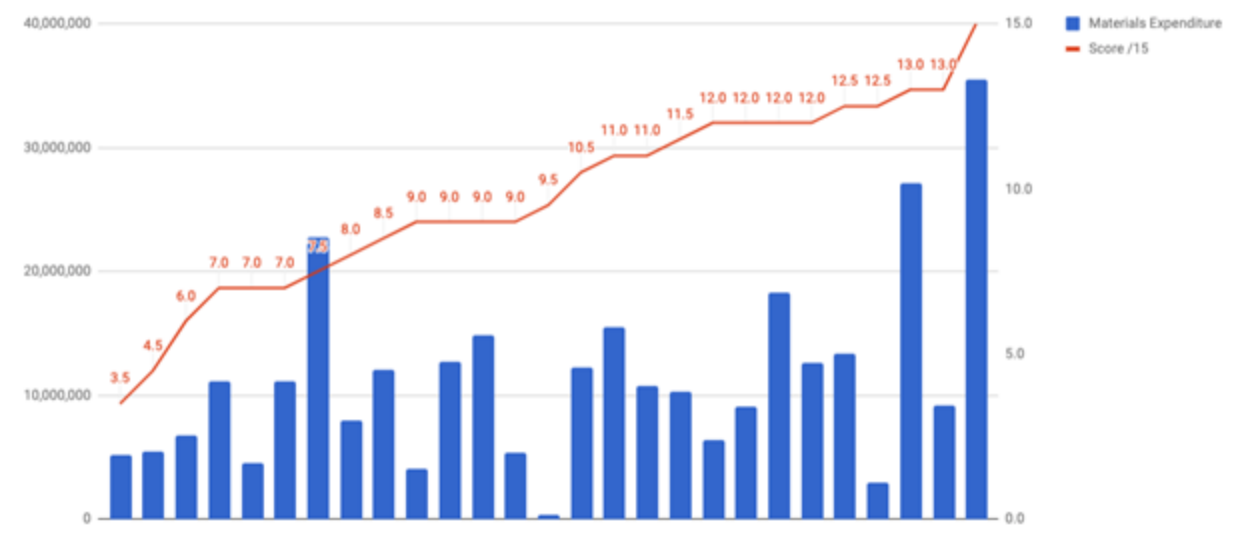
- CARL DPWG, Portage Network
- CRKN/Canadiana
- TAATU meetup (via the ACA)

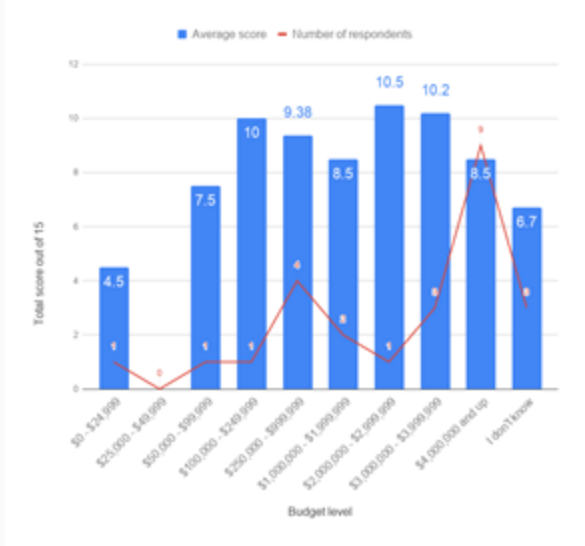
Regional (Canada)

- Library consortia: COPPUL (West), OCUL (Ontario), BCI (Quebec),, CAUL (Atlantic)
- Provincial associations with meetups, training opportunities: AAO, AABC, etc.

Capacity vs. resources

You don’t have to be a big, well-resourced institution to do digital preservation

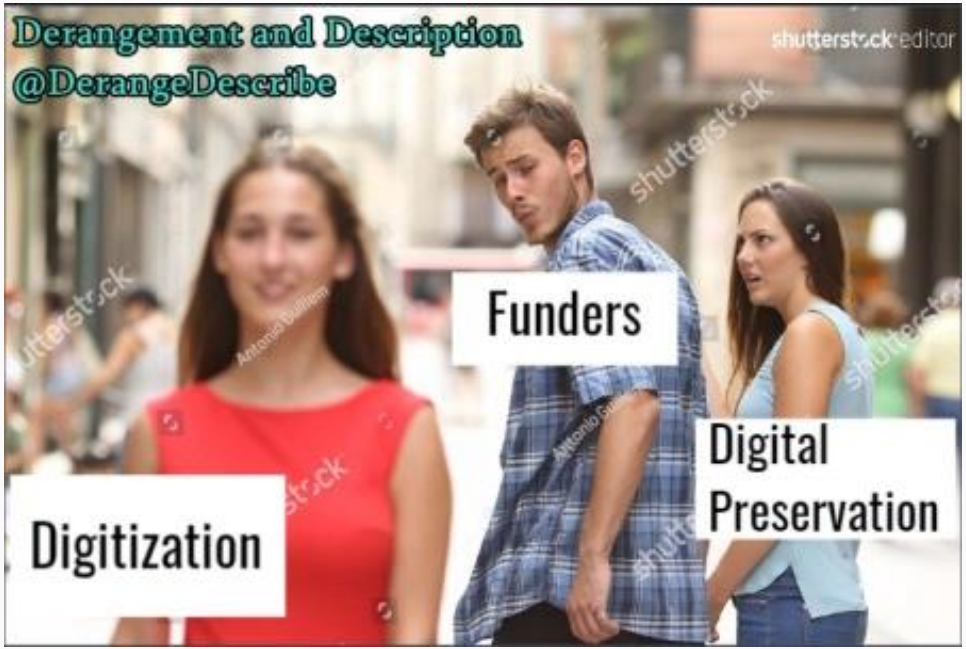




Overall score against materials expenditures, CARL members (left) and budget ranges, non-CARL (right)

Motivation

Funders are starting to require digital preservation plans (well, eventually)



So, what about the “who” and “where” of digital preservation?

- Current focus in academic library/archives sector is on how to share labour, costs for digital preservation work
- Developing models combine the traditional roles of memory institutions (working with donors, appraisal, mediating access) with collaborative services for the technical side (tools, infrastructure, training)
- Libraries/archives know their user community; service providers know the technology

Permafrost model

Service provider provides:

- Access to hosted tools for preservation tasks and maintains those tools (in this case, Archivematica)
- Access to hosted, scalable, cheap(er) storage infrastructure (OLRC)
- Training, documentation, advice
- Assistance with integrations with other systems, including for access

Memory institution:

- Decides what to preserve and which staff members will do the work
- Physically processes materials for preservation
- Applies institutional policies to preservation decisions
- Determines access requirements

Current challenges

- Figuring out fair and cost-effective ways to distribute digital preservation capacity, especially outside academic sector or working across sectors (with government archives at all levels, community archives, corporate archives)
- Lack of resources at the institutional level even if collaborative solutions are available
- Lack of consistently available training resources for different learning needs

Some nice readings

Furness, A., Marks, S., Robichaud, D., & Romkey, S. (2020). Doing digital preservation: Notes from the Digital Preservation Management Workshop public panel. *Off the Record 36*(1), 23-36.

Lyons, B. (2016, February 18). What is the chemistry of digital preservation? *AVPreserve*.

Walsh, T. (2017). How to access digital files from the nineties. *Canadian Centre for Architecture*.

Thank you!

Feel free to contact me at grant.hurley@utoronto.ca!