Key Issues Impacting Digital Preservation in Canada – 2021

An Executive Brief from the CARL Digital Preservation Working Group

This document was prepared by Steve Marks, Corey Davis, Grant Hurley, Rebecca Ross, Roslynn Ross, and Dale Askey.

November 2021

www.carl-abrc.ca

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES · ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Introduction

The challenge of digital preservation has been an evergreen topic in libraries for around two decades. Accordingly, the landscape now includes a number of public and private sector entities offering a range of infrastructure options. A subset of the CARL Digital Preservation Working Group (DPWG) offered to delve deeper into the question of infrastructure. Our first thought was an environmental scan, but the major actors are relatively well known. We also discussed preparing a recommendation or roadmap for creating a national Canadian digital preservation storage infrastructure, but that seems premature without broader consultation around some of the challenges and variables.

What follows frames digital preservation around four key drivers that we see at this moment:

- The emergence of the Digital Research Alliance of Canada,
- The future of Trusted Digital Repositories (TDRs),
- The state of digital preservation storage infrastructure in Canada, and
- The changing players on the field, particularly the entrance of commercial entities.

This document outlines the current state, opportunities, and risks inherent to each of these four areas. As you review these points, we recommend considering the following questions to help shape further investigation and actions.

- What do you want us to investigate further? Where do you need action?
- What do we need to know about the state of digital preservation in Canada?
- What is the landscape you want to see?
- Who Controls this Landscape?

Digital Research Alliance of Canada

Current State

- After years of consultation and planning, the federal government has inaugurated the Digital Research Alliance of Canada (formerly NDRIO), an organization that combines the previous mandates of Compute Canada, CFI, CARL's Portage project, and some aspects of CANARIE (the diagram halfway down this web page is instructive). This group is going to have high visibility in the research data management, or RDM, space (including its data preservation

aspects) due to its substantial initial funding and probably a mindshare boost from the recently released Tri-Agency RDM Policy.

- The Alliance has inherited, among others, the Portage Preservation Expert Group (PEG) and is keen to promote Alliance-supported services such as Dataverse and FRDR not only as access repositories but also as data preservation services.
  - You may remember that PEG co-wrote (with CARL's DPWG) a position paper (as a submission to the NDRIO consultation) highlighting what it saw as priorities and relevant players in the preservation of research data.
  - A PEG working group also produced a Preservation Recommendations Report for Scholars Portal Dataverse that involves connecting institutional subscriber Dataverse repositories to robust and replicated long-term storage of datasets in the Ontario Library Research Cloud (OLRC), ensuring a minimum level of reliable repository storage underlying Dataverse to support "preservation in place."

Opportunities

- We anticipate an increase in engagement around data management, including data preservation. Libraries were important in the development of the Alliance due to the CARL Portage initiative, so local stakeholders may be interested to hear what libraries have to say in campus discussions about RDM.

Risks

- These are mostly funding- and stability-related.
  - As a new organization, the Alliance's strength of mandate is not really known yet. Will its recommendations regarding data management and preservation be rigorously implemented by the federal granting agencies and other bodies?
    - Also related to its new status, it is unclear how this organization's funding and/or mandate will weather successive federal administrations. Depending on digital preservation services supported entirely or in part by the Alliance could bear additional risk until more is known.
  - The Alliance combines the mandates of multiple previous bodies. Will funding for data management and, within that, for preservation activities, be able to carve out space against funding demands for High Performance Computing?
  - What is the Alliance's long-term vision for data preservation in Canada? This has yet to be articulated clearly.

The Future of Trusted Digital Repositories (TDRs)

Current State
- Changes are afoot for Canada's two certified TDRs!
  - In 2018, the Canadian Research Knowledge Network (CRKN)'s TDR Task Group (TDRTG) made a recommendation to the CRKN Board to adopt the Scholars Portal TDR as a national TDR service for scholarly resources licensed by CRKN members, which was then endorsed by the membership at the 2018 CRKN AGM. CRKN continues to work closely with the University of Toronto and Scholars Portal on the development of an MOU that would enact the recommendations of the TDRTG and guide the next steps of this initiative. At the 2020 CRKN Conference, CRKN members re-confirmed their commitment and interest in this project.
  - CRKN has recently launched a Canadiana [TDR Audit Group](#) to undertake an [audit of the Canadiana TDR](#). In addition, a community-based audit group made up of CRKN members and stakeholders further increases Canadian capacity in the ongoing certification requirements of TDRs. Since the merger between Canadiana and CRKN in 2018, the Canadiana TDR is fully integrated as a CRKN service.
- At least partly because of [Portage/Alliance initiatives](#), [Core Trust Seal](#) is gaining visibility and community mindshare as an alternative to traditional ISO16363-based TDRs.
  - Although CTS comes from the data repository world (as a successor to the Data Seal of Approval and World Data System certification programs) it has the same philosophical underpinnings as TRAC/ISO16363-based TDR certifications.

Opportunities
- The conversation around TDR certification seems to be shifting from one where judgment is rendered by a central authority to one where consensus on the trustworthiness of a repository emerges from the community.
  - Does a Canadian community for TDR certification present an opportunity to promote specific standards and practices?
  - Can there be a standard/harmonized approach to things like repository contingency planning?

Risks
- Most of the repositories currently invested in certification are existing shared repositories. Stakeholders should take care to ensure that any conversations occurring in these groups reflect the actual needs of their institutions.

- TDR certification is an activity that requires a lot of time and resourcing. Is the focus on certification coming at a detriment to the actual work of doing preservation?

Digital Preservation Storage Infrastructure

Current state
- In Canada there are currently two main consortial storage providers: Scholars Portal's [Ontario Library Research Cloud](#) (OLRC) and COPPUL's [WestVault](#).
- Scholars Portal's OLRC is currently in a build-out phase via an infrastructure renewal project and is setting up test nodes at Canadian institutions outside of Ontario. Scholars Portal is also working on revised agreements and new pricing for non-OCUL universities, consortia, and other memory institutions.

Opportunities
- Are there opportunities for OLRC to harmonize service offerings across WestVault and other academic storage providers? Could different services exist for different communities but utilize a common technical back-end?
- If OLRC becomes a more widely used service, are there opportunities to integrate with platforms where data is being ingested and/or processed?
  - Scholars Portal's [Permafrost](#) service provides one example of how this might work. Scholars Portal has also been working on similar integrations with its Dataverse service.
- As always with commodity infrastructure, larger regional- or national-level collaborations can represent significant cost savings through leveraging economies of scale.

Risks
- The cost structure of locally administered storage services is often different from that of commercial cloud providers. A lack of understanding of this fact by the community could lead to decision-making that sets memory institutions up for nasty surprises like data retrieval fees when they need to take data out of commercial providers.
- Lack of decisive decision-making by infrastructure governance groups could lead to missed opportunities for adoption by Canadian institutions and strengthen the position of commercial and non-academic service providers.

A Changing Field of Players

Current State

- Internet Archive is again making moves to set up a larger presence in Canada. They are in the process of setting up a Canadian data centre (on the West Coast.) They are also currently in talks with the University of Victoria about a colocation arrangement that would see Internet Archive nodes located in the UVic data centre. They have also expressed an openness to colocated nodes at other institutions as well, though there don't seem to be any definite plans there. An MOU with the Alliance is in progress that would see IA acting as a Preservation Service Provider for FRDR.
- CRKN is in the early stages of exploring digital preservation services for Canadian institutions as well.
- Members of the CARL Digital Preservation Working Group are seeing increased chatter and activity from commercial vendors in this space, such as Preservica's free "Starter" offering and EBSCO's Perpetua service via Arkivum. Some of this is coming from their expansion into service offerings in the research lifecycle, some is just general questing tendrils.

Opportunities

- New services are often presented as "loss leader" offerings. If institutions can use these services while the costs remain low, and at the same time avoid pitfalls like vendor lock-in, there may be opportunities to minimize costs in the short term, at least temporarily.
- (Internet Archive specific) An increased IA presence in Canada could lead to new opportunities for integration with commonly used services such as Archive-IT and IA's digitization services.

Risks

- As always, outside organizations can be unpredictable in the ways that their service offerings and pricing structures develop over time. This can present dynamics that are inimical to the practice of digital preservation.
  - Need to be very cognizant of possible issues in the compatibility space, as vendor lock-in and proprietary formats are a well-established tactic.
  - Commercial services represent an opportunity to "outsource" preservation expertise, but expertise shed by libraries in such a way rarely comes back.
- (Internet Archive specific) Some within the CARL community view the ongoing lawsuit against IA around its National Emergency Library initiative during the COVID-19 pandemic with concern as it regards the ongoing sustainability of IA operations.

Conclusion

The goal with this briefing document is not to provide a comprehensive accounting or analysis of the state of digital preservation in Canada, but rather--as the title states-- to highlight some key issues where libraries need to make decisions individually and collectively in the coming years. As our need for digital preservation solutions increases, the inevitable growth of the commercial market and the influence of organizations such as the Alliance will require close observation to avoid taking paths that lead to a loss of control over our digital assets, spiraling costs, and/or a digital preservation failure. There is also growing awareness that we need to be cognizant that our strategies for some digital assets may not align with the needs and desires of Indigenous communities. These key notions and others raised in this document point back to the questions asked at the outset of this document:

- What do you want us to investigate further? Where do you need action?
- What do we need to know about the state of digital preservation in Canada?
- What is the landscape you want to see?
- Who controls this landscape?

If libraries and their directors can articulate clear answers to these questions, this will simplify the work by allowing those working in the digital preservation space to avoid some of the noise and focus their energies building the solutions we need in Canada. We have a real opportunity to create robust national infrastructure, services, and expertise for and around digital preservation if we can set a clear agenda and work collectively to achieve it.