



Library and Archives Canada: Towards a Trusted Digital Repository

Pam Armstrong
Library and Archives Canada
Ottawa, Canada

Meeting:

84. Preservation and Conservation, (PAC), Information Technology, IFLA-CDNL Alliance for Bibliographic Standards (ICABS) and Law Libraries

Simultaneous Interpretation:

Not available

WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL

10-14 August 2008, Québec, Canada
<http://www.ifla.org/IV/ifla74/index.htm>

Abstract

The Library and Archives Canada, legislated to be the enduring memory of the nation, is committed to becoming a Trusted Digital Repository, to provide reliable and long term access to the digital documentary heritage of Canada. Library and Archives Canada is working collaboratively with the Government of Canada, Canadian publishers, Canadian digital initiatives and the international domain. This paper will provide a corporate overview of the development of the institution's digital preservation infrastructure, an external view of collaboration with creators and national partners and a national strategic view of the implementation of a network of trusted digital repositories.

Introduction

In all domains the volume of digital information is increasing at an extraordinary rate, becoming the key driver of economic growth and social well being in this century. Our intellectual capital, found in educational, scientific, public, cultural and other intellectual resources, is increasingly at risk by the volatile character of digital objects and rapid developments in information technology. The strategies and tactics for preserving this information and providing long term access to it are being developed at various strata of collaboration and cooperation throughout the world. Library and Archives Canada has adopted a multi-pronged approach to adopting these strategies and is working at a corporate level, a federal government level, national and international levels to work through the requirements and issues affecting us all. This paper will provide a high level overview of the activities currently underway at Library and Archives Canada and in collaboration with Canadian digital publishers, government departments, memory institutions and organizations.

Library and Archives Canada Mandate

Library and Archives Canada is a newly formed knowledge institution, combining the collections, services and staff of the former National Library of Canada and National Archives of Canada. As the preamble of the 2004 Library and Archives of Canada Act states, the Library and Archives Canada is mandated to preserve the documentary heritage of Canada for the benefit of present and future generations, to be a source of enduring knowledge accessible to all, to facilitate cooperation in Canada among communities involved in the acquisition, preservation and diffusion of knowledge and to serve as the continuing memory of the government of Canada.¹

The legislation also carried with it very specific powers that would require the Library and Archives Canada to address its digital preservation capacity: Legal Deposit for electronic publications, the power to harvest the Internet for web content of interest to Canadians and the transfer of government or ministerial records considered to have historical or archival value or government records considered to be at risk. In addition to electronic monographs and serials, government records and websites, the Library and Archives Canada acquires digital photo and art, digital audio-visual, geomatics, electronic theses from Canadian universities, digital technical and architectural drawings, private textual electronic records and broadcast data content. As well, the Library and Archives Canada generates considerable volumes of digital content through its own large scale digitization program. These digital collections are acquired from a variety of domains, through a range of acquisition authorities and transfer protocols. They are managed with multiple metadata standards and storage environments and they are accessed through many different channels.

Library and Archives Canada Trusted Digital Repository

To address these legislated obligations and the preservation challenges inherent in these digital collections, Library and Archives Canada has embarked on a multi-year development project to establish a reliable, flexible, integrated digital preservation infrastructure and a suite of trusted digital repository services based on OAIS, the Open Archival Information System Reference Model, from which our digital documentary heritage can be identified, collected, managed, preserved and accessed.

From an organizational and administrative perspective, the Library and Archives Canada is laying the groundwork with the establishment of an executive level committee known as the Digital Information Steering Committee. Ensuring corporate collaboration and horizontal approaches to solutions, this group oversees and approves the work of all significant digital initiatives in the institution: digitization, digital acquisition, standards, digital preservation and digital access. New functional areas related to digital repository standards, government electronic records requirements, web archiving and electronic Legal Deposit have been established and staffed. Associated with these areas are new policies, including digital preservation, digital collection development and guidelines for

¹ Library and Archives of Canada Act <http://lois.justice.gc.ca/en/ShowDoc/cs/L-7.7>

website selection.² Further policies addressing digital file format guidance, digitization standards and electronic record transfer requirements are in development.

The technical underpinnings of trusted services will be achieved through the Library and Archives Canada Trusted Digital Repository Project. This is an implementation of a blend of software approaches; commercial, open source and custom built. Funded through a Government of Canada Treasury Board submission this development effort is largely accomplished by Library and Archives Canada staff and consultants working on collaborative projects with Canadian content creators; notably digital publishers and government departments. The project began with the development of ingest and Submission Information Package services for electronic legal deposit and government electronic archival records through the establishment of a Virtual Loading Dock.

The Virtual Loading Dock is the gateway to the Trusted Digital Repository, designed to receive and unpack assets, provide workflow functionality for curators and a tools framework which includes the following:

- MD5 (performs a checksum against the ingested assets for verification that asset manipulation does not alter the asset or lose data)
- McAfee (performs a virus scan of all ingested assets)
- JHOVE (physical asset format and integrity validation)
<http://hul.harvard.edu/jhove/>
- Interwoven MetaTagger 4.1.0 (Descriptive metadata extraction)
- DROID (automated batch identification of file formats - accesses the PRONOM DROID registry hosted as a service by the UK Archives)
<http://droid.sourceforge.net/wiki/index.php/Introduction>
- WinZip (decompression of packaged assets)

The following metadata standards are employed:

- METS (Metadata Encoding Transmission Schema)
<http://www.loc.gov/standards/mets>
- MODS (Metadata Object Description Schema)
<http://www.loc.gov/standards/mods>
- MARC (Machine Readable Cataloguing) <http://www.loc.gov/marc>
- PREMIS (Preservation Metadata: Implementation Strategies)
<http://www.loc.gov/standards/premis>
- METS Simple Rights Schema
<http://www.loc.gov/standards/mets/news080503.html>
- Dublin Core <http://dublincore.org>
- Government of Canada Records Management Metadata Standard <http://www.lac-bac.gc.ca/information-management/002/007002-5002.27-e.html>
- OAIS, Reference Model for an Open Archival Information System. CCSDS 650.0-B-1, Blue Book, January 2002
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

² Library and Archives Canada Digital Collection Development Policy
<http://www.collectionscanada.gc.ca/collection/003-200-e.html>

Work is currently underway to define the remaining architectural touch points to implement the archival store, the preservation masters, file format migration strategies and the access play-out services. The following diagram demonstrates the high level design for the Trusted Digital Repository, based on the Open Archival Information System Reference Model.

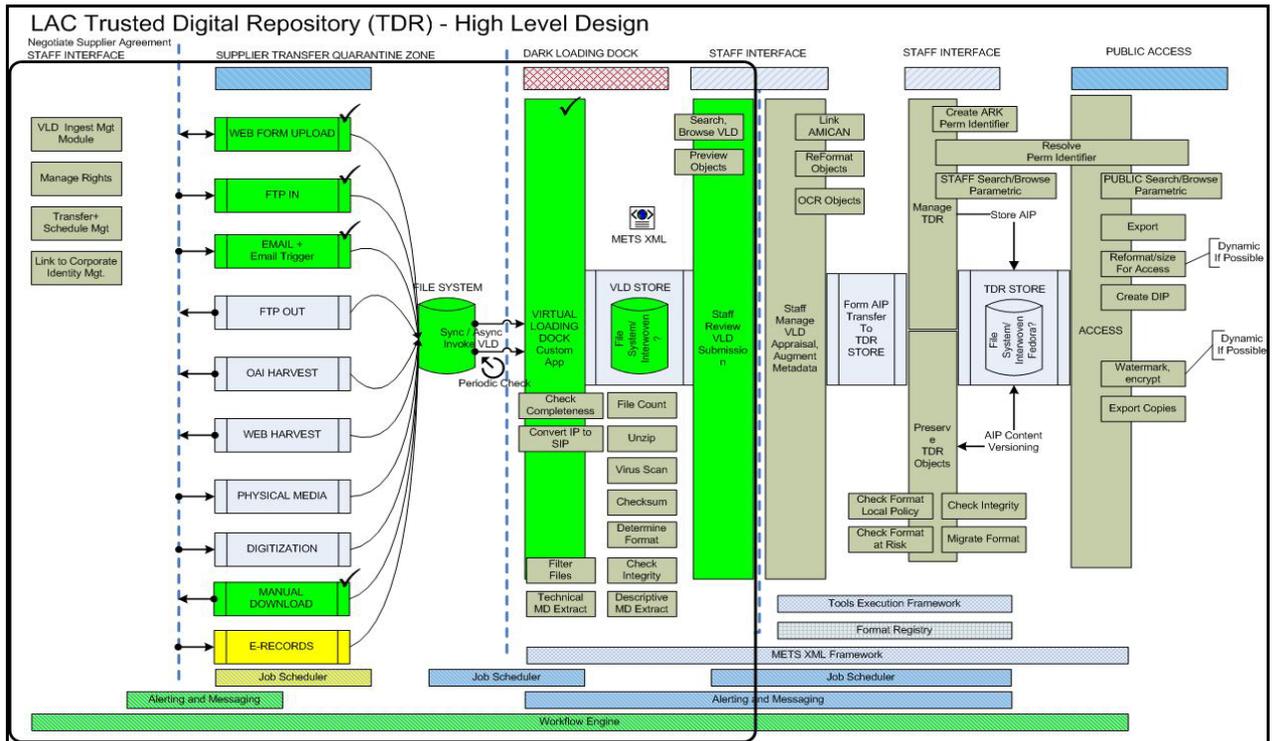


Figure 1. TDR High-Level Design

Collaboration with Government of Canada

Collaboration with content creators is a key component to the development of the Virtual Loading Dock. Within the federal government domain, Library and Archives Canada has initiated a pilot project in partnership with two government departments to transfer the electronic component of designated archival records under particular Record Disposition Authorities. Through this collaboration, Library and Archives Canada has developed an electronic record transfer tool which will be embedded in the Government of Canada Records, Document and Information Management System (RDIMS)³. This tool includes functionality to address search and selection of required records to be transferred, a metadata application profile for transfer, selection review and record packaging, secure transfer with Public Key Infrastructure, audit and transfer logs, integration with Library and Archives Canada Client Information Management system and the Library and Archives Canada Record Disposition Authority Control System. This particular ingest

³ RDIMS <http://www.tpsgc-pwgsc.gc.ca/rdims/text/about-e.html>

tool will be further developed as new government departments come on board, and development and operations of the transfer tool will be managed jointly with Public Works and Government Services Canada, the government department which provides common services to other federal departments.

Collaboration with Canadian Publishers

Collaboration is also underway with Canadian digital publishers. Library and Archives Canada has acquired digital publications over the last 10 years, although never before through legal deposit nor within an Open Archival Information System Reference Model framework. The Virtual Loading Dock will now capture key Submission Information Package metadata through management of the Publisher Profile which includes contact information, preferred ingest channels and schedules along with access permissions such as open, restricted, onsite-use only and embargoed. The Virtual Loading Dock has been launched with selected publishers from across Canada using only File Transfer Protocol, email, webform and manual ingest channels. We are working through their concerns and issues and at the same time we are discussing new harvesting techniques and further projects for collaboration. Library and Archives Canada has implemented a web archiving program using the open suite of tools developed by the Internet Archive and the International Internet Preservation Consortium. This suite of tools will also be implemented to serve as a Legal Deposit ingest channel.

Canadian Digital Information Strategy

While working at the corporate level with content creators from government and from the Canadian publishing domain, Library and Archives Canada is also working strategically at the national level with the creation of the Canadian Digital Information Strategy. In 2005, Library and Archives Canada initiated a dialogue reflecting a range of interests in the digital field with the goal of framing a national strategy. Through a series of meetings Library and Archives Canada consulted over 200 stakeholder organizations from a variety of sectors: publishing and media producers, creators, rights bodies, academics, provincial and federal officials and memory institutions. The consultations culminated in a National Summit at which broad consensus on the elements of a national strategy emerged. The proposed vision is as follows:

“Canada's digital information assets are created, managed and preserved to ensure that a significant Canadian digital presence and record is available to present and future generations, and that Canada's position in a global digital information economy is enhanced.”⁴

The three pillars of the vision include strengthening content creation, ensuring preservation of the content and maximizing access and use. Actions related primarily to digital preservation include improved production practices, selection and capture of digital content for long term retention, a distributed digital preservation repository

⁴ Canadian Digital Information Strategy <http://www.collectionscanada.gc.ca/cdis/index-e.html>

network, preservation related research, new workplace skills and increased public awareness of digital preservation issues.

Highlighting the need for continued collaboration the strategy maintains that it “cannot be undertaken by any single organization; rather, an inclusive, coordinated and distributed approach involving stakeholders from all sectors of the information environment is required. The information community sees growing urgency in the need to deal with digital issues, and has expressed a new willingness to work collaboratively within a common framework so that Canada is ensured a leadership role in the global digital knowledge economy.”⁵

National Initiatives

There are numerous initiatives within Canada which contribute to the goals of the Canadian Digital Information Strategy and it is important to highlight just a few of these activities. The InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems) at the University of British Columbia explores the preservation of authentic digital information.⁶ The Synergies Project, led by Université de Montréal, will make possible national accessibility to Canadian social sciences and humanities research.⁷ The Alouette Metadata Toolkit Project, led by the Canadian Association of Research Libraries provides open source software for simple digital content management and metadata management of digitized resources.⁸

Canada Project

The Canada Project is a partnership that includes Library and Archives Canada, Bibliothèque et archives nationales du Québec, Canadiana.org, University of Waterloo and the Canadian Association of Research Libraries. We are joining forces to encourage institutions, organizations and businesses across the country to pool their content to create a unique showcase of our nation’s heritage and a made-for-Canadians digital space to learn, innovate, and create. A cornerstone of the Canada Project is a commitment to adhere to the highest possible format and metadata standards and a commitment to long-term preservation of digitized and born digital content supported within a network of digital repositories, aggregators and Trusted Digital Repositories.

There are potentially thousands of small digital repositories across Canada holding unique collections. There are many and varied digital object repositories ranging from private collections to collections held and maintained by private and public institutions across Canada. The following diagram demonstrates the overview of data storage, aggregation and digital preservation.

⁵ Canadian Digital Information Strategy <http://www.collectionscanada.gc.ca/cdis/index-e.html>

⁶ InterPARES http://www.interpares.org/ip2/ip2_index.cfm

⁷ Synergies http://www.synergiescanada.org/index_en.html

⁸ AlouetteCanada Metadata Toolkit <http://www.alouettecanada.ca/nmrdi/nmrdi-e.php>

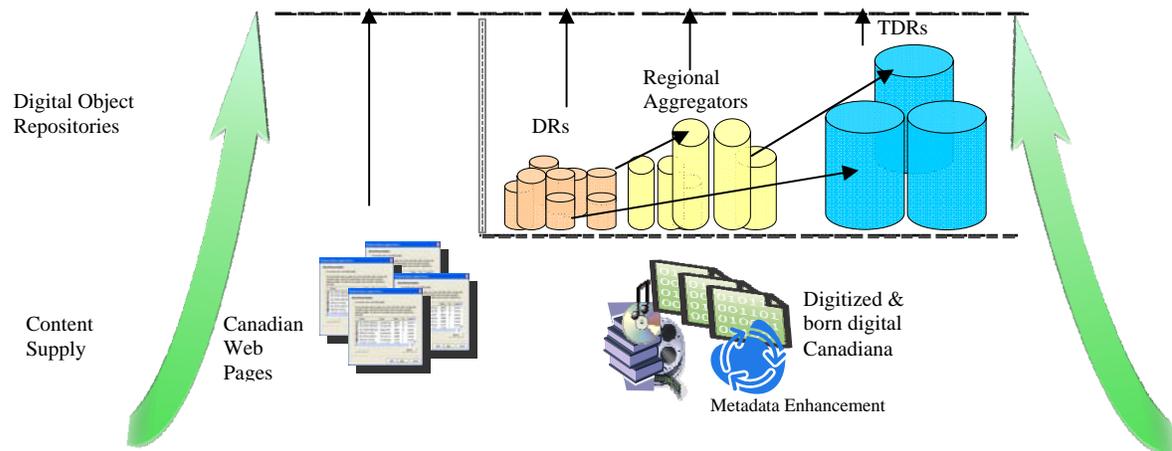


Figure 2. Digital Repositories

Aggregators are digital repositories for media types, subject disciplines, geographic or cultural content. There could also be metadata-only aggregators. Aggregators may provide a consolidation service through a common web interface or even shared server/storage facilities. We can anticipate numerous aggregators in larger cities and provincial capitals.

Trusted Digital Repositories provide the highest standards for the perpetual preservation of collections. Features include scalability, security, reliability, performance and continuity. In addition to the Library and Archives Canada Trusted Digital Repository currently under construction, several more Trusted Digital Repositories are foreseen being established in major centres across Canada. This network will provide capacity, performance and mutual redundancy and backup. Trusted Digital Repositories should be built on different technology platforms to provide as wide a span of experience possible to maximize the research value of the project as well as provide alternative long term paths for Trusted Digital Repository technology evolution.

To develop a national network of Trusted Digital Repositories there is much work to do. Library and Archives Canada is set to launch a Canadian digital preservation survey to establish a baseline of understanding digital preservation practice among Canadian repositories. This will include both large and small institutions, all types of content including digitized, born digital and data content: audio-visual, government, scientific, research and cultural domains. The survey will cover institutions on a targeted sampling basis. Based on this information and through negotiations with other institutions, Library and Archives Canada can then move forward on developing a collaborative national digital preservation infrastructure. Responsibilities for preservation will be shared through formal agreements or will originate within institutions according to their mandates.

International Collaboration

The international landscape of cooperation in the digital preservation sphere is invaluable to Library and Archives Canada. The tools, protocols and standards which we adopt in our own digital preservation infrastructures are all developed through international efforts and coded to industry best practices and standards. In addition to membership within such organizations as the International Internet Preservation Consortium, PREMIS and countless standards bodies, we have also developed informal networks with colleagues from around the world who share their knowledge and expertise most willingly. In this particular field it is extremely important to learn from each other, to advance the research and to cooperate where we can. This IFLA conference theme of “navigating towards global understanding” is a perfect expression of how together we can meet the challenges of ensuring continuing access to the world’s digital information resources.

Conclusion

This paper provides a high level overview of the development of the Trusted Digital Repository at the Library and Archives Canada. The Library and Archives Canada depends on collaboration and cooperation to achieve its mandate and works with the Government of Canada, Canadian publishers and national initiatives and organizations to advance the cause of digital information preservation in Canada. At the same time the Library and Archives Canada is very fortunate to work with international partners and colleagues to advance our work in Canada and where possible, to contribute to a global understanding.

Acknowledgements

Special thanks to my colleagues at Library and Archives Canada; particularly Susan Haigh, Manager, Digital Policy, Steve Sekerak, Enterprise Architect, Merideth Fletcher, Project Analyst and Marie-Claude Renaud, Technical Lead, Trusted Digital Repository Project.

References

1. Library and Archives of Canada Act <http://lois.justice.gc.ca/en/ShowDoc/cs/L-7.7>
2. Library and Archives Canada Digital Collection Development Policy <http://www.collectionscanada.gc.ca/collection/003-200-e.html>
3. RDIMS <http://www.tpsgc-pwgsc.gc.ca/rdims/text/about-e.html>
4. Canadian Digital Information Strategy <http://www.collectionscanada.gc.ca/cdis/index-e.html>
5. Canadian Digital Information Strategy <http://www.collectionscanada.gc.ca/cdis/index-e.html>
6. InterPARES http://www.interpares.org/ip2/ip2_index.cfm
7. Synergies http://www.synergiescanada.org/index_en.html
8. AlouetteCanada Metadata Toolkit <http://www.alouettecanada.ca/nmrdi/nmrdi-e.php>