

---

# Library and Archives Canada Towards a Trusted Digital Repository

October, 2008

Contact: [Pam.armstrong@lac-bac.gc.ca](mailto:Pam.armstrong@lac-bac.gc.ca)



Bibliothèque et Archives  
Canada

Library and Archives  
Canada

Canada 



# Introduction

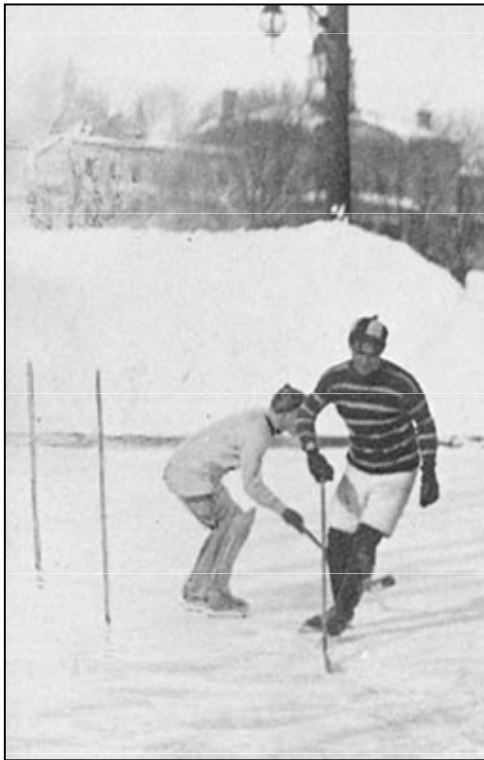
---

- ❖ Library and Archives Canada (LAC) is committed to providing reliable and long term access to the digital documentary heritage of Canada through development of trusted digital preservation services
- ❖ LAC is working collaboratively with partner institutions in the Government of Canada, Canadian publishers, Canadian digital initiatives and the international domain

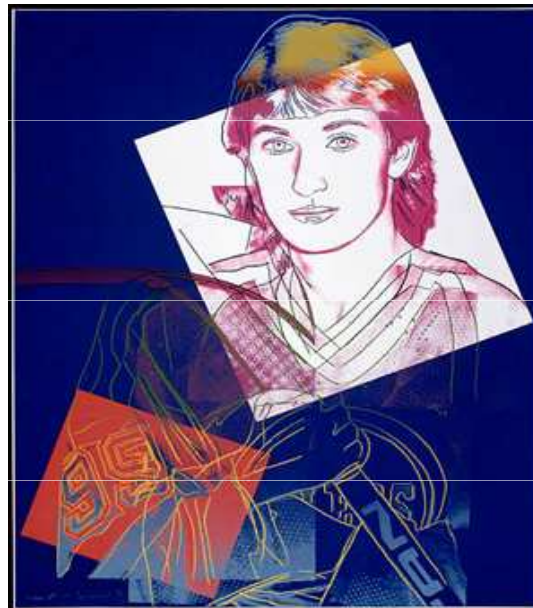
# Library and Archives Canada Mandate

---

**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 serve as the continuing memory of the government of Canada and its institutions.**



*Library and Archives Canada Act*



# Library and Archives Canada Act 2004

---

- ❖ A new knowledge institution combining the collections, services and staff of the former National Archives and National Library of Canada
- ❖ Legislated powers requiring the LAC to address its digital preservation capacity:
  - Legal Deposit of Electronic Collections
  - Transfer of federal government or ministerial digital records of archival value and federal government records at risk
  - Power to harvest the Internet for web content of interest to Canadians

# LAC Digital Collections – large, varied and growing

---



- ❖ Includes digital publications, websites, blogs, electronic government records, digital photo and art, digital audio-visual, geomatics, electronic theses from Canadian universities, digital technical and architectural drawings, private textual electronic records
- ❖ Managed via various standards, systems and procedures



# Trusted Digital Repositories : adopting a new digital preservation framework

---

- ❖ "A Trusted Digital Repository is one whose mission is to provide reliable and long-term access to managed digital resources to its designated community, now and in the future. "

<http://www.rlg.org/legacy/longterm/repositories.pdf>

- ❖ Attributes are financial, organizational, administrative, and technological
- ❖ LAC's TDR is based on the Open Archival Information System Reference Model (functions include Ingest, Data management, Archival storage, Preservation, Access)

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

# Addressing the challenge of preserving digital collections at LAC : developing trust through commitment

---

- ❖ **Financial commitment:** Established a multi-year project funded by Government of Canada Treasury Board to develop a suite of Trusted Digital Repository business and technology services. Collections in scope for this funding are the electronic publications and the Government electronic records.
- ❖ **Organizational commitment:** Addressing organizational issues with respect to corporate digital governance bodies and digital preservation functions
- ❖ **Policy commitment:** New policy development such as the
  - LAC Digital Collection Development Policy
  - Guidelines for Website Selection
  - Resource Description for Digital Publications Policy
  - LAC Digital Preservation Policy
- ❖ **Administrative commitment:** Internal and national standards development and adoption is underway: file formats, metadata, persistent identifier, rights
- ❖ **Technical commitment:** Development of Trusted Digital Repository and storage infrastructure

# Development of the LAC Trusted Digital Repository

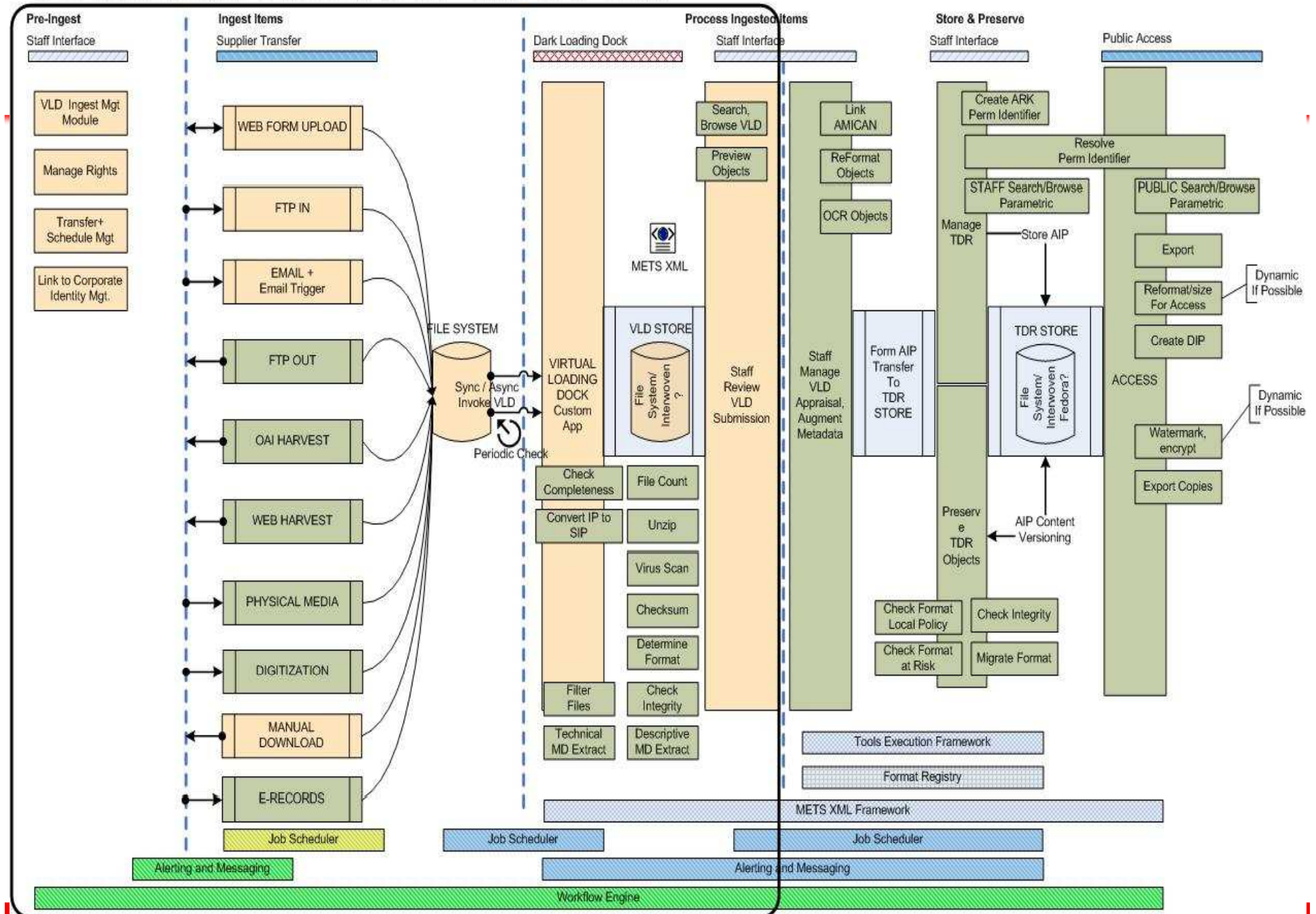
---

- ❖ The Trusted Digital Repository will be developed with generic functional modules, strong business workflows and collection-specific functionality
- ❖ Based on OAIS (Open Archives Information System Reference Model), we first developed the Ingest module which we call the Virtual Loading Dock (VLD)
- ❖ The Ingest module is the most important of all modules as actions and metadata captured in this phase enable long-term preservation
- ❖ The VLD is the gateway to the TDR, capturing all digital content acquired by LAC
- ❖ Assets are stored in the VLD until they are selected for permanent acquisition
- ❖ The VLD is designed to receive assets, validate their integrity, extract technical and descriptive metadata about the assets and prepare the OAIS Submission Information Package





# LAC Trusted Digital Repository (TDR) - High Level Design



# LAC TDR Metadata Standards

---

- **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.oclc.org/research/projects/pmwg/>
- **Government of Canada Records Management Metadata Standard**
- **Dublin Core** <http://dublincore.org/>

# LAC Open Protocols and Tools

---

- ❖ **OAIS Open Archive Information System Reference Model**
- ❖ **OAI – Open Archive Initiative metadata harvester**
- ❖ **JHOVE – Metadata extract + checksum creation**
- ❖ **DROID – Digital object recognition**
- ❖ **Pronom – File format registry**
- ❖ **Heretrix/Wayback – Web harvesting, search, view**
- ❖ **ARK – Persistent Identifier minting**
- ❖ **LDAP – Authentication and role management**
- ❖ **PureFTP, Secure FTP– supports LAC security zone**
- ❖ **HTTP Apache, Tomcat, PHP**
- ❖ **SOAP + REST- Web services**
- ❖ **Search protocols (Z39.50, OpenURL ...)**

# Collaboration with Government of Canada Partners

---

- ❖ Completing first phase of a pilot project in partnership with Treasury Board Secretariat to transfer the electronic component of designated archival records under certain Record Disposition Authorities
- ❖ Developed a transfer application to be embedded in the Government of Canada “Records, Document and Information Management System”
- ❖ Completed first phase by transferring several thousand electronic records to the LAC
  - VIA Secure Channel, Secure File Transfer Protocol
  - Transferred records and associated metadata
  - Final records and versions, multiple file formats

# Collaboration with Canadian Publishers

---

- ❖ LAC has acquired electronic publications over the previous 10 years, never before within an OAIS context, never before through legal deposit
- ❖ Virtual Loading Dock now captures key Submission Information including the publisher profile through online registration, publication profile details and technical, structural and rights metadata
- ❖ METS and PREMIS have served as excellent information models to ensure the proper capture of metadata
- ❖ Launched Virtual Loading Dock with selected Canadian publishers who are working with us to identify performance and functional issues
- ❖ Supports
  - FTP
  - Email
  - Web form
  - Manual ingest



# Depositing at the Virtual Loading Dock

The image displays a sequence of four overlapping screenshots from the Library and Archives Canada website, illustrating the steps for depositing a publication at the Virtual Loading Dock.

**Screenshot 1: Publisher Details**

Library and Archives Canada / Bibliothèque et Archives Canada

**Canada**

Français	Contact Us	Help	Search	Canada Site
Home	About Us	What's New	What's On	Publications

**Publisher Details**

Publisher Name: Paul Saker  
Place of Publication: Kanata  
Publisher URL: <http://www.psake>  
Access Rights: ALL

**Publication File Types**

What type of publication would you like to deposit?

- ☒ Monograph
- ☐ New Serial
- ☐ Serial Issue
- ☐ Website

**Screenshot 2: Access Permissions**

Library and Archives Canada / Bibliothèque et Archives Canada

**Canada**

Français	Contact Us	Help	Search
Home	About Us	What's New	What's On

**Access Permissions**

Current Access: [ ]  
If access permissions for this publication differ from the default, please specify:  
Embargo: Starting [ ]

**Screenshot 3: Publication Details**

Library and Archives Canada / Bibliothèque et Archives Canada

**Canada**

Français	Contact Us	Help	Search	Canada Site
Home	About Us	What's New	What's On	Publications

**Publication Details**

ISBN: [ ]  
ISSN: [ ]

**Screenshot 4: File Details**

Library and Archives Canada / Bibliothèque et Archives Canada

**Canada**

Français	Contact Us	Help
Home	About Us	What's New

**File Details**

File to Upload: [ ]



# Web Archiving at LAC

---

- ❖ Began Government of Canada domain harvesting in 2005
- ❖ Other harvested websites include Provincial/Territorial domains, Canadian Olympic websites, Federal Election 2006 and related political websites
- ❖ Use open source web harvesting software, Heritrix, viewing software, Wayback
- ❖ LAC developed our Dashboard to manage crawls, indexing and quality control. The Dashboard will be available to Open source community this fiscal year.
- ❖ LAC has completed 3 GC domain complete harvests from 2005-2008.
- ❖ Public access to Government of Canada Web Archive via the Internet has been available since Nov. 2007.
- ❖ Full text searchable, browse by government department.
- ❖ LAC is member of IIPC Steering Committee, Digital Preservation and Access Working Groups.

# Preservation Features of LAC TDR

---

- ❖ LAC Digital Preservation Policy completed spring 2008 and predicated on a file format migration strategy, strong metadata and replication
- ❖ Working on preservation implementation strategies
  - Enhanced metadata capture with PREMIS framework (technical, structural, rights, provenance, descriptive)
  - Metadata guidelines for creators
  - File format guidelines for creators
  - File format migration strategies (on ingest and based on risk assessment)
  - Maintain original and migrate forward
  - Using TNA PRONOM and DROID for digital format registry and format identification
  - Embrace replication locally, nationally , internationally
  - Technology watch for emulation and new technologies



# Access features of LAC TDR

---

- ❖ Implementing rights schema based on PREMIS
  - Open
  - “Moving wall” = Embargoed access
  - Restricted (closed)
  - On-site access only
  - Copyright
- ❖ Implementing persistent identifier schema called ARK
- ❖ Creating service copies for public access, respecting accessibility issues for citizens

# Where we are at!

---

- ❖ VLD was soft-launched for legal deposit of electronic publications on May 2<sup>nd</sup>, 2008
- ❖ Currently defining the architecture and identifying the technology solutions for the archival store of the TDR
- ❖ Currently defining the requirements for Government Records ingest & archival services
- ❖ Currently looking at the integration of mass digitization processes to the VLD/TDR
- ❖ On target for end-of-fiscal delivery:
  - Core functionality for TDR archival storage
  - Persistent Identification for TDR assets
  - Streamlining technical infrastructure
- ❖ The preservation and access components of the TDR will be addressed in FY09-10; the foundation is currently being established

# Challenges

---

- ❖ New frontier; forging a new path, there are no TDR models to follow
- ❖ New business paradigm, blending IT and business
- ❖ Ensuring a balance between TDR requirements and gaining business efficiencies for staff and end-users
- ❖ Treasury Board funding covers only portion of LAC digital acquisition and holdings