

Metadata and the Digital Ingest Workflow

Considerations from the 2019 Proof of Concept (PoC) with Public Services and Procurement Canada (PSPC) on the receipt of EDRMS records from OpenText to Library and Archives Canada (LAC)

Presentation to the DCMI Virtual 2020 Best Practices
2020-09-24

Library and Archives Canada



Library and Archives
Canada

Bibliothèque et Archives
Canada

Canada

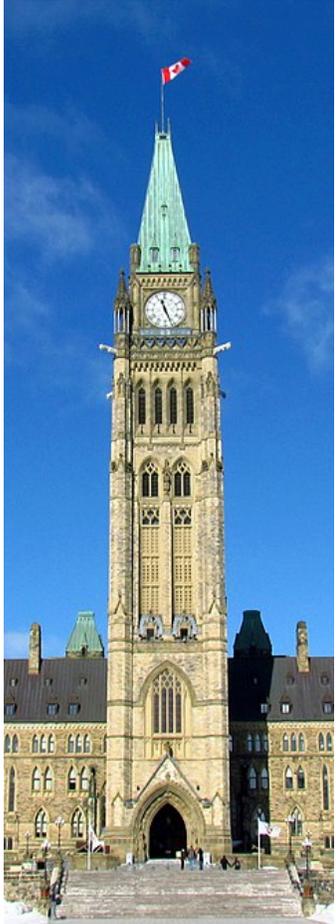
Today's Menu

1. What was LAC's approach to the project?
2. What resources did we have available?
3. What did we test, and what did we find?
4. What challenges, lessons, and considerations do we have?
5. How will we move forward?



**Why would we
ever want to
take on this
challenge?**

Policy Context / Project Purpose



Disposition of Archival Records (April 1, 2017)

As of April 1, 2017, Library & Archives Canada (LAC) requires that Government of Canada (GC) Institutions, subject to the LAC Act, transfer digital formatted archival records to LAC in *digital format*.

Directive on Service and Digital (April 1, 2020)

Supersedes Directive on Recordkeeping, Policy on IM, and related policies on April 1, 2020.

Key points:

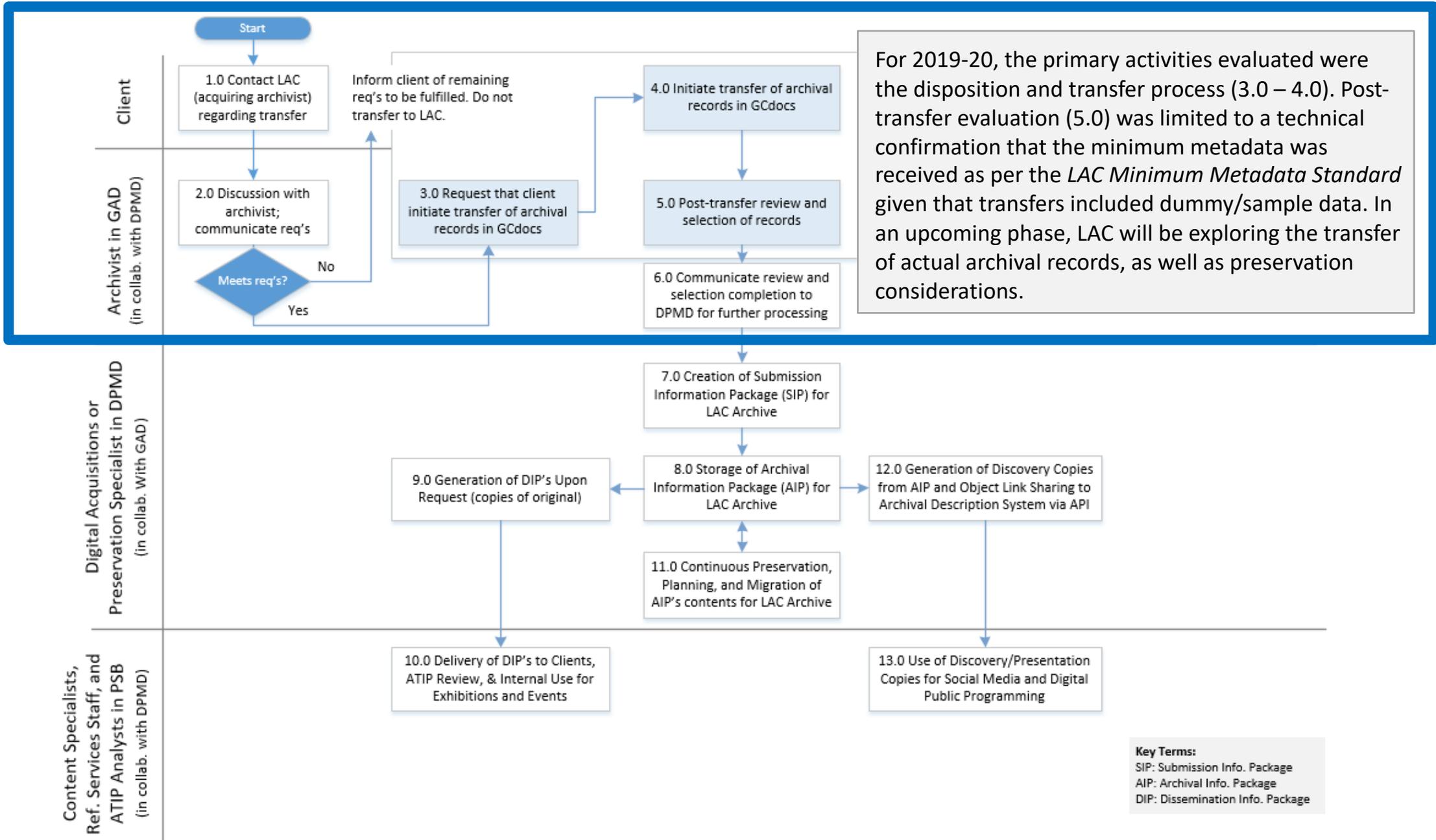
- Requires that departments enable interoperability, reuse, and sharing of information to the greatest extent possible within *and with other departments* across the government;
- Requires that departments use an approved GC enterprise IM solution to document business activities and decisions (e.g. GCdocs);
- Requires that departments *develop a documented disposition process and perform regular disposition activities for all information and data, as required*;

Objective and Intent



- The **objective** of this pilot project was to test that a client department using the Government of Canada's recommended EDRMS (GCdocs) could (i) create a disposition to archive a set of objects (i.e. documents, records) using automated retention triggers, and then (ii) transfer that disposition to Library and Archives Canada (LAC) for analysis and archive fulfillment.
- Specifically, we wanted to **validate the feasibility of applying the LAC minimum metadata standard to GCdocs**, a standard that included both descriptive metadata and technical (preservation) metadata requirements.
- Performing **preservation actions** on the data is a future activity for testing, though 2019-20 provided insights for future consideration.

Proposed Future State GCdocs Workflow for Testing: Digital Transfer Process for GCdocs – AGR GCdocs PoC 2019-20



What was LAC's approach to the project? (phases and deliverables)

Q1 2019-20

Q2 2019-20

Q3 2019-20

Phase I : Define the Requirements

✓  PoC Requirements
Traceability Matrix (Excel)

Phase II: Define Future State Based on Requirements

✓  Draft Future State
Process Map (Visio)

✓  Final Future State
Process Map (Visio)

Phase III: Define Test of Future State and Execute Tests

✓  Draft User
Acceptance Test
(UAT) Script

✓  Draft User Acceptance
Test (UAT) Script

✓  UAT testing and test results (PPT/Word)

✓  PoC/Workflow
Initial Build

✓  PoC/Workflow Build
(iterative updates)

✓  Project Findings and
Recommendations (PPT)

What was LAC's approach to the project? (it takes an army...)

Role	Position	Contact
Project Authorities / Managers	Consulting Project Executive, GCDocs Program	Lloyd Hayes
	Senior Project Officer, Office of Deputy Librarian and Archivist of Canada, LAC	Kyle Browness
Project Functional Lead (PSPC)	Project Oversight & GCDocs Program Functional Lead	Patrick Plouffe
Project Functional Lead (LAC)	Manager, Recordkeeping Strategies	Marie-Claude Côté
Technical Lead (PSPC)	Technical Analyst, Digital Services Branch, PSPC	David Willson
Technical Lead (LAC)	Team Lead, Innovation and Digital Transformation, LAC	Sylvain Marcil
Technical Consultant (LAC)	Business Consultant, Innovation and Digital Transformation, LAC	Vinod Chettur
Business Analyst / Technical Writer	Business Analyst (Consultant), Digital Services Branch, Innov. & Integration, GC-Wide Programs and Solutions, PSPC	Hugh Sheldon
Subject Matter Experts (PSPC) (Government Recordkeeping)	Senior Product Officer, GCdocs Program GC-Wide Programs and Solutions, Digital Services Branch, PSPC	Emilie Ouellet
	Senior Product Officer, GCdocs Program GC-Wide Programs and Solutions, Digital Services Branch, PSPC	Mary Wilde
Subject Matter Experts (LAC) (Government Recordkeeping)	Senior Project Officer, Government Records and Initiatives Division, LAC	Bonnie Clark
	Senior Project Officer, Government Records and Initiatives Division, LAC	Marissa Paron
	Senior Project Manager, Government Records and Initiatives Division, LAC	Joelle Dagenais
	Project Officer, Digital Integration, LAC	Heather Tompkins
Test Client Archivists	Archivist, Archives Branch, LAC (for Natural Resources Canada)	Martha Sellens
	Archivist, Archives Branch, LAC (for Atlantic Canada Opportunities Agency)	Adam Coombs
	Senior Archivist, Archives Branch, LAC (for Statistics Canada)	Jennifer Cuffe
	Lead Archivist, Archives Branch, LAC (for Office of the Commissioner of Lobbying)	Jenna Smith

What was LAC's approach to the project? (it takes good governance...)



Role	Position	Contact
PSPC/LAC Governance	Director General, Archives Branch, LAC (Project Sponsor)	Robert McIntosh
	Director General, Digital Operations and Preservation Branch, LAC	Sylvain Bélanger
	Senior Director General, Innovation and Chief Information Officer Branch, LAC	Surinder Komal (Q1-Q2) Dominique Bouvier (Q3-Q4)
	Director of Innovation and Digital Transformation, LAC	Simon Pageot
	Director of Digital Preservation and Migration, LAC	Roslynn Ross
	Director of Government Archives Division, LAC	Sandy Ramos
	Director of Government Records and Initiatives Division, LAC	Candace Loewen
	Director of Enterprise Project Management Office, PSPC	Jennifer Woods
	Director of IT Strategy and GC Interop Solutions, PSPC	Phillippe Lefebvre
	Director of GC-Wide Programs and Solutions, PSPC	Megan Hnatiw
Test Client Members and Stakeholder Governance	A/Manager, IM Policy and Planning, Natural Resources Canada	Martin Legault
	Information Management Analyst, IM Division, Natural Resources Canada	Meghan Batho
	Information Administrator, IM Division, Natural Resources Canada	Lisa Grierson
	A/Manager, Information Management Policy and Planning, Natural Resources Canada	Christine Newman
	Director, Information Management, Atlantic Canada Opportunities Agency	William Bembridge
	Manager, Information Management, Chief Information Officer Directorate, Atlantic Canada Opportunities Agency	Sophie Doiron
	IM Analyst, CIO Directorate, Atlantic Canada Opportunities Agency	Jennifer Cook
	IM Analyst, CIO Directorate, Atlantic Canada Opportunities Agency	Marie Mannette
	Chief, GCdocs Project, IM Division, Statistics Canada	Margaret Devey
	Head, Recordkeeping, Document Management Center, Statistics Canada	Randall Myles
Chief Information Officer, Office of the Commissioner of Lobbying Canada	Kevin Trudeau	

What resources or case studies did we have available to plan and implement the work?

Library and Archives Canada Minimum Metadata Standard for Transfers of Digital Archival Government Records

1. Effective Date

This standard was approved by the Director General, Archives Branch, Operations Sector, and came into effect on XXXX, 2020.

2. Application

This standard applies to all government records in digital format considered to have historical or archival value (hereinafter referred to as "digital archival government records") that are disposed of by means of transfer to the care and control of Library and Archives Canada (LAC) per section 13 of the *Library and Archives of Canada Act*. It is intended for use by information management (IM) and recordkeeping (RK) professionals in government institutions who are responsible for the disposition of government records, and recommended for those who configure the metadata infrastructure of software solutions for managing digital government records.

This standard must be met by government institutions at the time of disposition, in concert with all applicable legal instruments and other LAC policies governing the transfer of digital archival government records. While it outlines the baseline of metadata that is mandatory for all digital archival government records transferred to LAC, additional metadata requirements apply to supplement this standard for the following categories of records:

- Non-textual (e.g., graphic; audio-visual) and multi-media records;
- Structured information (e.g., data sets from scientific databases); and
- Digitized versions of analogue source records.

The scope of application of this standard is limited to metadata that describe archival records themselves; a separate instrument will be used to communicate metadata requirements for describing the package used to transfer this content to LAC. Specifications for the format and structure according to which the mandatory minimum metadata are to be provided to LAC at the time of transfer fall outside of the scope of this instrument, but will be the focus of a complementary policy tool.

3. Definitions

Definitions to be used in the interpretation of this standard are included in [Appendix A](#).

Minimum Metadata Standard for Transfers of Digital Archival Government Records

- DRAFT, not yet published; LAC will make available more widely soon;
- Provided the metadata requirements to support effective testing;
- Includes both **descriptive** metadata requirements (e.g. title, language, etc.) and **technical** (preservation) metadata requirements, such as:
 - **Integrity (item level)**: Information used to confirm that the contents and format of the resource described have not been altered at the bit level.
 - **Provenance (item level)**: Details about [originating] hardware, [media] or system components that are pertinent to rendering the resource(s) accurately over time.

Draft Minimum Metadata Standard (key categories)

Metadata Concept	Definition	Data Mapped for GCdocs PoC
Format	The specific structure or arrangement of data code stored as a computer file whose purpose is to dictate how a computer displays, prints, processes and saves it.	<i>File Properties</i> section, e.g. MIME Type (e.g. File / PDF v. 4.2, etc.).
Extent	The size or duration (dimensions) of the content of the resource.	<i>Extent</i> (pages, words)
Language	A language of the intellectual content of the resource.	<i>Lang</i> attributes for title
Resource Identifier	A unique value or code that is applied to the resource in a given system or context.	<i>Identifier</i> , e.g. 154665
Title	A name given to the resource.	<i>Title</i> (eng; fre)
Date/Time	A point or period of time associated with an event in the lifecycle of the resource (e.g., an action taken on the resource, such as creation).	<i>Record Creation Date</i> <i>Record Change Date</i> (last modified)
Creator	An agent – i.e., the individual or group (organizational unit or corporate body) – primarily responsible for making (i.e., creating, accumulating and/or maintaining) the resource.	<i>Affiliation: Agent Name, Agent Corporate Name, Agent Section</i>
Classification Code	A unique code or identifier derived from a recordkeeping file plan that is applied to the resource.	<i>Classification</i> , e.g. GC-1-010-01

Draft Minimum Metadata Standard

(key categories cont.)

Metadata Concept	Definition	Field(s) Mapped for GCdocs PoC
Rights Mngt. Information	Permissions, limitations or restrictions to accessing or using the resource, including if and when a record may be partially or fully released (e.g., declassified or opened) for public access or use, or formally published.	Security Clearance > <i>Current Security Clearance</i> . Stored as <i>Access Condition, Use and Reproduction, Copyright</i>
Integrity	Information used to confirm that the contents and format of the resource described have not been altered at the bit level.	Checksum workflow-generated prior to transfer to LAC (MD5)
Technical Environment [or Provenance] Information	Details about [originating] hardware, [media] or system components that are pertinent to rendering the resource(s) accurately over time.	<i>File Properties</i> section, e.g. MIME Type (e.g. File / PDF; Adobe Acrobat v. 2.1, etc.)
Disposition Authority	In accordance with the provisions of the Library and Archives of Canada Act, the instrument that the Librarian and Archivist of Canada issues to enable government institutions to dispose of records which no longer have operational utility, either by permitting their destruction, by requiring their transfer to LAC or by agreeing to their alienation from the control of the GC.	Entered when creating or editing a file classification, e.g. INST-2019-001 <i>Record Info Note > Disposition Authority</i>

What we tested, and what we found



The EDRMS content archiving workflow should, among a long list of requirements:

1. Send **object-level metadata in an open standards-based XML format** (non-proprietary). Each XML is associated clearly with its associated object. (achieved: MODS XML 3.7; required export tool; XML with each object)
2. Include **all descriptive and technical metadata** as per *LAC's Minimum Metadata Standard*. This includes a checksum generated on the file prior to being transferred to LAC. (achieved)
3. Password **protections and encryptions must be removed** by the Producer prior to transfer. (user-dependent; can automate detection but not removal)
4. Workflow **performance tested** with files of at least 1 GB each, and single transfers of at least 10,000 files, diacritics/special characters, and to confirm folder paths of 260+ characters can be processed. (achieved)
5. Original **arrangement structure must be included** (for LAC, having arrangement in metadata was sufficient as folders and sub-folders are virtual in the EDRMS whereas all data is held in a 'flat' structure) (achieved)

What we tested, and what we found (metadata standard can be effectively applied)

Element Name	Definition	notes	example	Encoding scheme	Standard on
Access Restrictions	Permissions assigned to an information resource that govern or restrict the ability of users to locate or take actions on the resource.	This element is applied to the resource (not the individual). It protects the resource from unauthorized access or alteration. "Access restrictions" relate to who may locate or take actions on the resource. "Usage Conditions" relate to the use of the content of the resource.	Read-only Full access	Institution-specific encoding scheme	Mandatory
Agent Identifier	A unique value that identifies the individual who performs an action on an information resource.	Typically a coded user name such as Smith2	Idea75	locb	
Agent Name	The name of an individual who performs an action on an information resource.	Actual name, not coded identifier.	Doe, Jane	Institu encod	
Agent Corporate Name	The name of an individual who performs an action on an information resource.		resources Canada	Regis	
Agent Section Name	The name of an individual who performs an action on an information resource.		Research Branch	Institu encod	
Agent Role	The role of an individual who performs an action on an information resource.		ator ber	Institu encod	

Standard on Metadata for Transferring Digital Archival Records from Government of Canada Institutions to Library and Archives Canada

1. Effective Date

These requirements were approved by the Director General, Archives Branch, Chief Operating Officer Sector, and came into effect on XXXX.

2. Application

This standard derives from sections 7(c), 12 and 13 of the *Library and Archives of Canada Act* (LAC Act), and the *Library and Archives Canada (LAC) Evaluation and Acquisition Policy Framework*. It defines the metadata that must accompany (i.e., that is mandatory for) all government records in digital format identified as being of historical or archival value ("digital archival government records" hereafter; alternatively referred to as digital "unpublished information resources of enduring value") that are transferred to the care and control of LAC. It likewise applies to records identified as archival within bilateral transfer agreements reached between LAC and Government of Canada (GC) institutions not subject to the LAC Act pursuant to sections 8(f) and 9 of the LAC Act. As such, this standard is to be met in concert with all applicable LAC policy and related instruments regarding transfer in force at the time of application.

This standard is to be used by information management (IM) and recordkeeping (RK) professionals within GC institutions who are responsible for disposing of digital archival government records by means of transfer to LAC. Ideally it will inform and be integrated into information technology (IT) and IM planning and development by GC institutions when devising systems to manage their digital records.

The metadata requirements outlined herein apply irrespective of the media type (e.g., graphic; audio-visual; textual) of the records in question. However, additional metadata requirements for the following types of digital records must be anticipated, as will be addressed in supplementary resources:

- Non-textual and multi-media records;

```

</role>
<!-- Agent Identifier -->
<nameIdentifier>
  <name>
    <!-- Language -->
    <language displayLabel="Language / Langue">
      <languageTerm type="code" authority="iso639-2b"/>
    </language>
    <!-- For now, we will park Integrity and Aggregation here -->
    <physicalDescription displayLabel="Physical description / Description physique">
      <form authority="marcform">electronic</form>
      <internetMediaType/>
      <extent/>
      <!-- Values of <digitalOrigin> are born digital, digitized microfilm, digitized other analog, reformatted digital -->
      <digitalOrigin displayLabel="born digital"/>
      <!-- Aggregation -->
      <note displayLabel="Aggregation Level"/>
      <!-- Integrity -->
      <note displayLabel="MD5 Checksum"/>
    </physicalDescription>
    <!-- Classification Code-->
    <classification displayLabel="Classification / Classification" authority="RM Classification"/>
    <!-- Resource Identifier -->
    <identifier displayLabel="GC System Identifier" type="GC System Identifier"/>
    <!-- Access Restrictions -->
    <accessCondition displayLabel="Access condition(s) / Condition(s) d'accès" type="restrictionsOnAccess"/>
    <!-- Record Information and changes -->
    <recordInfo displayLabel="Record information / Information(s) sur la notice">
      <recordContentSource authority="oclcorg">NLC</recordContentSource>
      <!-- Event Date/Time when Event Type == 'Created' -->
      <recordCreationDate/>
      <!-- Event Date/Time when Event Type not 'Final' or 'Created' -->
      <recordChangeDate xsi:type="dateOtherDefinition"/>
      <!-- Event Date/Time when Event Type == 'Final' -->
      <recordChangeDate/>
      <recordOrigin/>
      <languageOfCataloging displayLabel="Language of cataloguing / Langue de cataloguage">
        <languageTerm type="code" authority="iso639-2b"/>
      </languageOfCataloging>
      <descriptionStandard/>
    </recordInfo>
  </nameIdentifier>
</mods>

```

- GCdocs field elements re-used in the metadata to accompany dept. transfers to LAC;
- LAC's Minimum Metadata Transfer Requirements aligned with the GCdocs common set of metadata categories;
- Transfer mechanism leverages existing retention schedules and export functions.

What we tested, and what we found

(data can be exported and processed in our systems ... but more work to come)

opentext™ | Content Server

Enterprise Personal Tools My Account Search Search Enterprise

Enterprise > OpenTest >

OpenText Company History

OpenText Company History
Owner: [Browness, Kyle](#)

A history of the OpenText company [Edit description](#)

- Modified on 2019-10-21 13:24
- [You](#) added the most recent version at 2019-10-21 13:24

Size: 61 KB
Unrated
[Rate this item](#)

[Open in Desktop](#)
[Download](#)
[E-mail this link](#)

[Edit in Desktop](#)
[Add version](#)

[Notify me of changes](#)
[View properties](#)
[More actions](#)

Preservica

Kyle Browness English

Dashboard Ingest Access Preservation Data Management Administration Help

Root GCDOCS Testing PSPC_Export_ PSPC_Export_ OpenText Com...

Properties Advanced History XML

Title	OpenText Company History		
Description	A history of the OpenText company		
Type	Asset	Security Tag	open
Ref	413a5ffc-2bea-40a8-a383-97968baab54c		

Identifiers

This record has no identifiers

[Add Identifier](#)

Generic Metadata

mods, Namespace: <http://www.loc.gov/mods/v3>

schema Location	http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mo
title Info	
authority	iso639-2b
display Label	OpenText Company History
lang	English

Links

Parent	PSPC_Export_Oct172019	View Properties
--------	-----------------------	---------------------------------

Please select a template [add Fragment](#) [Save](#) [Cancel](#)

preservica.com Copyright © 2019 Preservica

Export and transfer of object(s) + XML metadata



Challenges, Lessons, and Considerations for EDRMS Archiving

Versioning. An EDRMS is a double-edged sword. It stores thousands of versions (*users: yay!*) ... but ... it stores thousands of versions.

Consideration:

You likely do not want the EDRMS to by-default export all versions of a document; if you do, ensure this is a strategic decision, or that the workflow can shift depending on your needs.

Sending all versions is technically easy but poses significant challenges for archival processing, costs for preservation, and public access issues.

Challenges, Lessons, and Considerations for EDRMS Archiving

Security. EDRMS security features are only as good as their application.

Consideration:

Be extra-cautious to ensure you feel confident that the records labelled as 'unclassified' are indeed so; archivists can do spot-checks, in addition to written confirmations from records providers at more senior levels of the organization.

Ensure that the system-specific categories of security for the EDRMS (e.g. 'open', 'closed', 'private') can be mapped to the archive's various channels for accepting unclassified vs. higher levels of classification.



```

</role>
  <!-- Agent Identifier -->
  <nameIdentifier/>
</name>
  <!-- Language -->
<language displayLabel="Language / Langue">
  <languageTerm type="code" authority="iso639-2b"/>
</language>
  <!-- For now, we will park Integrity and Aggregation here -->
<physicalDescription displayLabel="Physical description / Description physique">
  <form authority="marcform">electronic</form>
  <internetMediaType/>
  <extent/>
    <!-- Values of <digitalOrigin> are born digital, digitized microfilm, digitized otl
  <digitalOrigin>born digital</digitalOrigin>
    <!-- Aggregation -->
  <note displayLabel="Aggregation Level"/>
    <!-- Integrity -->
  <note displayLabel="MD5 Checksum"/>
</physicalDescription>
  <!-- Classification Code-->
<classification displayLabel="Classification / Classification" authority="RM Classif
  <!-- Resource Identifier -->
<identifier displayLabel="GC System Identifier" type="GC System Identifier"/>
  <!-- Access Restrictions -->
<accessCondition displayLabel="Access condition(s) / Condition(s) d'accès" type
  <!-- Record Information and changes -->
<recordInfo displayLabel="Record information / Information(s) sur la notice">
  <recordContentSource authority="oclcorg">NLC</recordContentSource>
    <!-- Event Date/Time when Event Type == 'Created' -->
  <recordCreationDate/>
    <!-- Event Date/Time when Event Type not 'Final' or 'Created'-->
  <recordChangeDate xsi:type="dateOtherDefinition"/>
    <!-- Event Date/Time when Event Type == 'Final' -->
  <recordChangeDate/>
  <recordOrigin/>
- <languageOfCataloging displayLabel="Language of cataloguing / Langue de c
  <languageTerm type="code" authority="iso639-2b"/>
</languageOfCataloging>
  <descriptionStandard/>
</recordInfo>
ods>

```

Challenges, Lessons, and Considerations for EDRMS Archiving

Preservation Metadata. Is it sufficient? Where are the gaps, and how will you fill them?

Consideration:

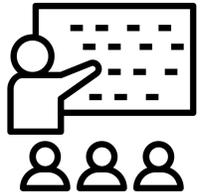
Does the expected information you receive from your records provider allow or inhibit you to perform all necessary preservation actions?

If you don't yet have a minimum technical metadata standard, implement one and refine over time.

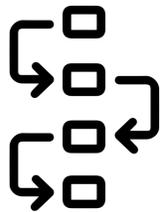
What about provenance?

What about data integrity?

How will we move forward?



1. Training and Outreach. LAC is working with the central body providing the GC EDRMS to provide more online training to record managers (via an online portal) and to develop outreach that clarifies Records Management responsibilities in government (IM Backstage Pass Days, twice annually) and disposition usage in the EDRMS (walkthrough video online).



2. Testing and Refinement of Workflow. LAC is working with its partner institutions in government to test transfers of archival records with more varied metadata (above and beyond minimum metadata standard). This includes testing automated file format migrations post-transfer and automated file-format filtering at-transfer to support sustainable access.



3. Security Refinement. The current workflow is certified for lower-privacy documents. LAC is working with security agencies in government to raise the security of the tools used for transfer to include more sensitive materials.

Marie-Claude Côté
Manager, Recordkeeping Strategies (on leave)
Archives Branch
mc.cote@canada.ca (mc.cote@me.com)
+1 (613) 979-0705

Library and Archives Canada

www.bac-lac.gc.ca

Telephone: 613-996-5115 or 1-866-578-7777 | TTY: 613-992-6969 or 1-866-299-1699

National
Capital Region

Vancouver
British Columbia

Winnipeg
Manitoba

Halifax
Nova Scotia



Library and Archives
Canada

Bibliothèque et Archives
Canada

Canada