Teamwork in Digital Stewardship: Moving Beyond Paper to Discovery

Placing Canadian Sessional Papers (1901-1925) into the Gov Info Repository

Sam-chin Li
Graham Stewart
Andrea Mills
What are Sessional Papers

- A collected series of reports, documents and papers tabled in the House or Senate
  - Annual Reports
  - Estimates and Public Accounts
  - Royal Commission and Task Force reports
  - Election Results
  - Government responses to Committee Reports
  - Government Policies
Rich Collection of primary resources for early Canadian history, economic, political science, etc....

But are all hidden in the Sessional Papers Collection
Example: insurance history in Canada
Example: Summary of the Trade of Canada with Principal Countries, 1868 - 1922
How can we make this valuable resource discoverable?
Many Sessional Papers were not printed, check the table of contents of each volume for details.

**Keywords:** Canada -- Politics and government 1867- Periodicals

**Downloads:** 105
<table>
<thead>
<tr>
<th>Title</th>
<th>Department</th>
<th>Content Date</th>
<th>Serial Title</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report of the Auditor General for the year ended 30th June, 1900. A-P</td>
<td>Auditor General’s Office</td>
<td>1899-1900</td>
<td>Audit Report</td>
<td>McDw</td>
</tr>
<tr>
<td>Report of the Auditor General for the year ended 30th June, 1900. Q-Y</td>
<td>Auditor General’s Office</td>
<td>1899-1900</td>
<td>Audit Report</td>
<td>McDw</td>
</tr>
<tr>
<td>Public Accounts of Canada, for the fiscal year ended 30th June, 1900.</td>
<td>Department of Finance</td>
<td>1899-1900</td>
<td>Public Accounts</td>
<td>Fiv W.S.--C</td>
</tr>
<tr>
<td>Estimates of sums required for the service of Canada, for the year ending on the 30th June, 1902.</td>
<td>Department of Finance</td>
<td>1901-1902</td>
<td>Estimates</td>
<td></td>
</tr>
<tr>
<td>Supplementary Estimates for the year ending 30th June, 1901.</td>
<td>Department of Finance</td>
<td>1900-1901</td>
<td>Estimates</td>
<td></td>
</tr>
<tr>
<td>Further Supplementary Estimates for the year ending 30th June, 1901.</td>
<td>Department of Finance</td>
<td>1901-1902</td>
<td>Estimates</td>
<td></td>
</tr>
</tbody>
</table>
Islandora is a possible Solution
Why move Sessional Papers to the Gov Info Repository using Islandora?
Digital asset management

Islandora
Tux is © lewing@isc.tamu.edu and the GNU Image Manipulation Program
(http://en.wikipedia.org/wiki/GIMP)
Apache web server
MySQL / PostgreSQL
PHP / Python
Tomcat

Linux

Tux is © lewing@isc.tamu.edu and the GNU Image Manipulation Program (http://en.wikipedia.org/wiki/GIMP)
Java

Apache web server
MySQL / PostgreSQL
PHP / Python
Tomcat

Linux

Tux is © lewing@isc.tamu.edu and the GNU Image Manipulation Program
(http://en.wikipedia.org/wiki/GIMP)
Apache web server
MySQL / PostgreSQL
PHP / Python
Tomcat
Linux

Drupal
FedoraCommons
Apache Solr

Java

Tux is © lewing@isc.tamu.edu and the GNU Image Manipulation Program (http://en.wikipedia.org/wiki/GIMP)
Apache web server
MySQL / PostgreSQL
PHP / Python
Tomcat

Drupal
FedoraCommons
Apache Solr

Java

Linux

Tux is © lewing@isc.tamu.edu and the GNU Image Manipulation Program (http://en.wikipedia.org/wiki/GIMP)
Stewardship

Islandora
Collaboration

Islandora
Access

Islandora
Flexibility

Islandora
Everything is an object

Source: https://wiki.duraspace.org/display/ISLANDORA711/Chapter+3+-+Getting+Started+with+Islandora
Objects have datastreams

Source: https://wiki.duraspace.org/display/ISLANDORA711/Chapter+3+-+Getting+Started+with+Islandora
Solution Packs

- Audio
- Basic image
- Large image
- Book

- Newspaper
- Video
- PDF
- Paged content
Content ingested \(\Rightarrow\) Datastreams created

- Tiff files (zipped)
- Metadata (MODS or MARCXML)
- Fedora object relationship
- MODS record
- DC record
- Thumbnail
Book Solution Pack

Datastreams created per page

- Binary object
- Technical metadata
- PDF
- JPEG 2000
- Medium JPEG

- OCR
- HOOCR
- DC record
- Thumbnail
- Fedora relationship data

Page object data
OpenSeadragon viewer for individual pages
Internet Archive bookreader for entire book
Full text searchable via Apache Solr
Ingestion

Workflow Overview

Metadata Files

Zipped Content

Islandora
What we have:

- JP2000 images
- Scanned and named in bound volumes
- Volume level metadata

What we need:

- TIFF images
- Organized and named by individual Sessional Paper
- Paper level metadata

Challenges
Workflow (1): Metadata Preparation

Create metadata for each Sessional Paper in a shared spreadsheet

Write individual marcxml files based on the spreadsheet
Workflow (2): Content Preparation

1. Isolate Sessional Paper
2. Transclusion
3. Conversion
**Transclusion: What is it?**

- *Transclusion* is the use of part or the whole of one or more items (e.g., text or images) in the creation of a new digital item.
- Relevant images are copied, renamed, derived and made available at a new URL.
- Internet Archive Tool.

**Workflow (2a): Content Preparation**
Transclusion

Workflow (2b): Content Preparation

Assess IA Scanned item: Determine Starts and Stops

Create Identifiers for each Sessional Paper

Transclusion Split each Sessional Paper
Files Conversion

Workflow (2c): Content Preparation

Download jp2 → Unzip → Convert tiff → Zip
Ingestion

Workflow (3)

Metadata Files

Zipped Content

Islandora
Interface Design

Workflow (4)
Government Information Repository
Government Information Repository: Sessional Papers
Gordon Belray  
(Information Architect)  
Sam-chin Li  
(Government Publications Librarian)  
Andrea Mills  
(Collection Specialist at Internet Archive Canada)  
Graham Stewart  
(Network Services Manager)  
Marlene Van Ballegooie  
(Metadata Librarian)  
Nicholas Worby  
(Graduate Student Library Assistant)