Perspectives on Preservation Using Islandora

Marcus Barnes and Mark Jordan
iCampEU 2019
Digital Preservation and Islandora

Policy
Preservation policies articulate your institutional commitment to making content accessible in the future.

Action
Action plans describe the tools, procedures, workflows, and outcomes used to implement your policies.

Auditing
Auditing enables you to demonstrate your preservation actions are achieving the commitments articulated in your preservation policies.
Simon Fraser University

Integrity checking
Islandora Checksum Checker

Executed every second day [MJ to confirm]

500 objects per execution

OBJ and MODS datastreams are checked

Takes about a month on our main Islandora repo
<table>
<thead>
<tr>
<th>AUDIT RECORD ID</th>
<th>DATE</th>
<th>EVENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUDREC8</td>
<td>2018-02-17T04:32:18.029Z</td>
<td>PREMIS:eventOutcome-SHA-1 checksum validated.</td>
</tr>
<tr>
<td>AUDREC10</td>
<td>2018-04-18T07:11:06.608Z</td>
<td>PREMIS:eventOutcome-SHA-1 checksum validated.</td>
</tr>
<tr>
<td>AUDREC11</td>
<td>2018-06-17T06:35:10.427Z</td>
<td>PREMIS:eventOutcome-SHA-1 checksum validated.</td>
</tr>
<tr>
<td>AUDREC17</td>
<td>2019-01-05T05:34:00.973Z</td>
<td>PREMIS:eventOutcome-SHA-1 checksum validated.</td>
</tr>
<tr>
<td>AUDREC18</td>
<td>2019-01-05T05:34:01.696Z</td>
<td>PREMIS:eventOutcome-SHA-1 checksum validated.</td>
</tr>
</tbody>
</table>
## PREMIS EVENTS

<table>
<thead>
<tr>
<th>FIELD</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>eventIdentifierType</td>
<td>Internal</td>
</tr>
<tr>
<td>eventIdentifierValue</td>
<td>MODS:AUDREC7</td>
</tr>
<tr>
<td>eventType</td>
<td>fixity check</td>
</tr>
<tr>
<td>eventDateTime</td>
<td>2018-02-17T04:32:16.840Z</td>
</tr>
<tr>
<td>eventOutcome</td>
<td>SHA-1 checksum validated.</td>
</tr>
<tr>
<td>eventIdentifierType</td>
<td>Internal</td>
</tr>
<tr>
<td>eventIdentifierValue</td>
<td>MODS:AUDREC9</td>
</tr>
<tr>
<td>eventType</td>
<td>fixity check</td>
</tr>
<tr>
<td>eventDateTime</td>
<td>2018-04-18T07:11:04.922Z</td>
</tr>
<tr>
<td>eventOutcome</td>
<td>SHA-1 checksum validated.</td>
</tr>
</tbody>
</table>
University of Toronto Scarborough

NDSA Levels of Preservation
Digital Scholarship Unit Preservation Strategy

- Guided by the foundational and operating principles of the University of Toronto Libraries’ Digital Preservation Policy.

- Practices are built based on the NDSA Levels of Digital Preservation recommendations.
Categories

Systems are in place to protect, know, monitor and repair data in five major categories:

1. Storage and Geographic Location
2. File Fixity and Data Integrity
3. Information Security
4. Metadata
5. File Formats
Storage and Geographic Location

NDSA recommendation: three complete copies of data are kept in separate geographic locations with different disaster threats.

1. Copy with Islandora (production servers with IITS managed system backups)
2. Separate copy to DSU managed system on North campus (snapshots and AIPs)
3. Archival Information Packages (generated using Islandora BagIt) exported to separate campus ITS DAMS (Project Canopus), preserved as part of UTL’s overall preservation strategy.
File Fixity and Data Integrity

1. Checksums generated on ingest *
2. Islandora Checksum Checker for testing data at scheduled intervals, as well as on demand.
3. Fixity information is indexed and visible through the DSU’s reporting framework.
4. Once AIPs have been transmitted to UTL’s central ITS department, unit staff no longer have write access to the Bags (no one person has write access to all copies).

(*) Fedora 5 includes the ability to provide transmission fixity.
DSU’s reporting framework provides practices that deal with many of the NDSA metadata recommendations.

In addition to advising best practices for data creation, the DSU performs transcoding to open formats.
AIP Creation

- Uses MODS XML `<reformattingQuality>` element to flag content ready for inclusion in preservation pipeline.
- Indexing specific events from Fedora’s AUDIT datastream as extended by Islandora modules that author specific PREMIS events.
- At regular intervals, use appropriate DSU Islandora reports scripts to get PIDs of objects for AIP creation. (By comparing indexed events.)
- AIPs are created using Drush commands from Islandora BagIt (and possible other tooling).
- AIPs are backed up and stored appropriately.
Preservation Calendar

- Regular server backups
- Monitor integrity of created Bags (AIP)
- Diff bags in North Campus against repository objects (every 6 months to a year)
- Data review (through the DSU’s Islandora reports scripts) and new bag generation.
- File Format review (for normalization) - generate a report (every 6 months to a year). Islandora Datastream CRUD can be used as part of normalization tasks.
Arca

Deposit into the WestVault distributed preservation storage network
WestVault

Drag and drop Files Here to Upload

Or Select Files to Upload
Islandora Westvault

Extends the Islandora BagIt module

Provides user interface for selecting objects to deposit into WestVault

Scheduled job to generate Bags for newly selected objects

Pushes to WestVault using the OwnCloud command-line client
'Checking this box creates a PRESERVATION datastream, which will flag this object for preservation on the next cron run. Un-checking this box will delete the PRESERVATION datastream, which will not delete the preserved objects, but stop the preservation of updates to the object.'
'New children of this collection (excludes current children)'

'All children of this collection (existing and new)'

'Current children only - new objects will not be preserved'
Drush commands

30 0 *** cd /var/www/drupal/sites/site1.myrepo.ca && drush -u 1 westvault-bagit

30 10 *** cd /var/www/drupal/sites/site1.myrepo.ca && drush -u 1 westvault-sync
Auditing in Arca

Confirm Bag was created

Confirm Bag contains the desired content

Only once confirmed should the PRESERVATION datastream be updated

WestVault also provides tools for determining what Bags have been deposited.
Riprap

Fixity microservice to replace Islandora Checksum Checker

Islandora integration module

Can handle fixity checks migrated from Islandora 7.x
<table>
<thead>
<tr>
<th>NAME</th>
<th>MEDIA TYPE</th>
<th>MIME TYPE</th>
<th>MEDIA USE</th>
<th>CHANGED</th>
<th>FIXITY AUDITING</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>NAME</th>
<th>MEDIA TYPE</th>
<th>MIME TYPE</th>
<th>MEDIA USE</th>
<th>CHANGED</th>
<th>FIXITY AUDITING</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMG_5081.jpg</td>
<td>Image</td>
<td>image/jpeg</td>
<td>Original File</td>
<td>Sun, 03/19/2019 - 13:23</td>
<td><a href="http://localhost:8080/mcrepo/res/2019-03/IMG_5081.jpg">Details</a></td>
</tr>
</tbody>
</table>
Islandora Bagger

Microservice to create Bags

Uses Islandora’s REST interface to get content

Command-line application with Drupal end-user integration

  Flexible BagIt options via integration with Context

  User clicks on “Create Bag” button, gets email with download link
Preservation Normalization Microservices

Don’t exist yet, but Islandora 8’s use of microservices sets the pattern.

Examples include:

- .doc to PDF
- Excel to CSV
- PDF to PDF/A

Potential to use Parcore Preservation Action Registries