

Preservation

Preservation File Formats / Normalization

Solution Packs follow preservation best practices

- Archival / Preservation Formats
- Ingested files can be normalized to a preferred archival format
 - Typical workflow would be to convert a PDF to PDF/a, a .docx to .odt, etc.
- Normalize to preservation and access formats

Fixity and File Format Identification

Checksum

- Enables checksum generation on datastreams
- Retrospective checksum generation

FITS

- Applied to canonical archival file (OBJ)
- File Format Identification
- Alternate Checksum Generation
- Extraction of Technical Metadata stored in a separate datastream.

Data Integrity

Checksum Checker

- Cron or user configured checksum verification
- Verification is recorded as a PREMIS-compatible event
- Alerts automatically set when mismatch discovered.

Bitstream/Object Replication

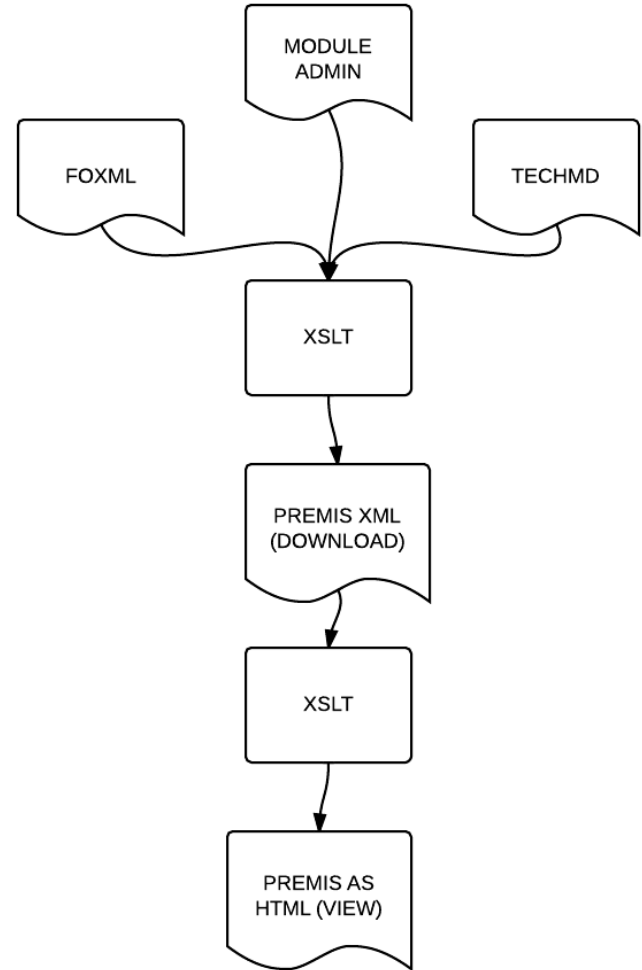
BagIt

- Bitstream/Object Copying
- manifest includes checksums
- Highly configurable
- Can include PREMIS XML
- Can produce Archivematica transfer Bags

Preservation Metadata

PREMIS

- Generates PREMIS XML on demand from Fedora Object's Audit Log, extracted technical, descriptive, and rights metadata
- Includes Agent, Event, Object and Rights entities



"City of Dover" : bought by Penetang group

View Manage MARCXML PREMIS

Download PREMIS

PREMIS Object

Field	Value
objectIdentifierType	Fedora Commons datastreamVersion ID
objectIdentifierValue	AUDIT.0
compositionLevel	0
messageDigestAlgorithm	
messageDigest	text/xml
formatName	Fedora Commons contentLocation REF value
contentLocationType	
objectIdentifierType	Fedora Commons datastreamVersion ID
objectIdentifierValue	RELS-EXT.0
compositionLevel	0
messageDigestAlgorithm	SHA-1
messageDigest	8acc007d964a3bf29e44d0978c1369051a6abbd1
size	544
formatName	application/rdf+xml
contentLocationType	Fedora Commons contentLocation REF value
objectIdentifierType	Fedora Commons datastreamVersion ID
objectIdentifierValue	MODS.0
compositionLevel	0
messageDigestAlgorithm	SHA-1
messageDigest	a94e53eb3f379cdd43594ae5504758340a08bbb
size	2387
formatName	text/xml
contentLocationType	Fedora Commons contentLocation REF value
objectIdentifierType	Fedora Commons datastreamVersion ID
objectIdentifierValue	DC.0
compositionLevel	0
messageDigestAlgorithm	SHA-1
messageDigest	ad931b32519134be6074e22da8f33279268585
size	1276
formatName	text/xml
contentLocationType	Fedora Commons contentLocation REF value
objectIdentifierType	Fedora Commons datastreamVersion ID
objectIdentifierValue	OBJ.0
compositionLevel	0
messageDigestAlgorithm	SHA-1
messageDigest	ca62da6b4ec9ea85c140295102cefabb72b4d
size	16129943
formatName	image/tiff
contentLocationType	Fedora Commons contentLocation REF value

PREMIS Events

Field	Value
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC5
event Type	fixity check
eventDateTime	2013-12-31T08:02:52.959Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC6
event Type	fixity check
eventDateTime	2013-12-31T08:02:53.058Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC7
event Type	fixity check
eventDateTime	2013-12-31T08:02:53.659Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC8
event Type	fixity check
eventDateTime	2014-01-11T11:04:59.993Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC9
event Type	fixity check
eventDateTime	2014-01-11T11:05:00.069Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC10
event Type	fixity check
eventDateTime	2014-01-11T11:05:00.177Z
eventOutcome	SHA-1 checksum validated.
eventIdentifierType	Internal
eventIdentifierValue	AUDIT.0:AUDREC11
event Type	fixity check
eventDateTime	2014-01-11T11:05:00.572Z
eventOutcome	SHA-1 checksum validated.

PREMIS Agent

Field	Value
agentIdentifierType	URI
agentIdentifierValue	http://library.yorku.ca
agentName	York University Library
agentType	organization
agentIdentifierType	URI
agentIdentifierValue	http://www.fedora-commons.org/
agentName	Fedora Repository 3.6.2
agentType	software

PREMIS Rights

Field	Value
rightsExtension	Photograph now in public domain. For citation information or to obtain high resolution image, contact ascpj@yorku.ca or see Toronto Telegram FAQ.

Preservation Interest Group

“The purpose of the [Islandora Preservation Interest Group](#) is to develop and communicate a common approach to address preservation within the Islandora framework.”

Check out the repo!

Islandora / Islandora-Preservation-Interest-Group

Unwatch 30

branch: master

Islandora-Preservation-Interest-Group / background_services_discussion_paper

Edit

32 commits

1 branch

0 releases

3 contributors

branch: master Islandora-Preservation-Interest-Group / +

couple of little formatting tweaks		
ruebot authored 8 days ago	latest commit 53da2eae8d	
background_services_discussion_paper	couple of little formatting tweaks	8 days ago
meetings	Meeting 01 notes	2 months ago
README.md	Remove Robin.	26 days ago

README.md

Islandora Preservation Interest Group

Convenors

- Donald Moses (UPEI)
- Nick Ruest (York)
- Mark Jordan (SFU)

Terms of Reference

- The name of the IIG is the Islandora Preservation Interest Group
- The purpose of the Islandora Preservation Interest Group is to develop and communicate a common

couple of little formatting tweaks

ruebot authored 8 days ago

..

README.md couple of little formatting tweaks

README.md

Islandora Background Processes Document

Mark Jordan, July 21, 2014

Overview and purpose of this document

This document proposes an alternative to the way that most derivatives are currently created by Islandora. It describes a generalized framework for Islandora that can accommodate both external (e.g., REST) processes running on the Drupal server. The proposed alternative can coexist with currently available solution packs, and microservice-based approaches to creating derivatives, if the solution packs are modified slightly.

The document also introduces a proof-of-concept implementation that demonstrates using a RESTful Optical Character Recognition service to create Islandora datastreams.

The purpose of the document is to encourage discussion around how the Islandora community, and communities supporting other repository platforms, can support a healthy ecosystem of services that can be shared across institutions, that require zero (or near-zero) configuration, and that are highly scalable.

Derivative creation in Islandora

Islandora solution packs typically create derivatives from the OBJ datastream as it is uploaded by the user in the target collection's ingest form. Uploading of the user's file, creating the derivatives, and submitting the ingest form are performed in sequence. Solution

Membership

- Mark Jordan (Simon Fraser University)
- Nick Ruest (York University)
- Donald Moses (UPEI)
- Melissa Anez (Islandora Foundation)
- Mark Leggott (UPEI)
- Ryan Gjerde (Luther College)
- Nathan Books (Northern Illinois University)
- Aaron Collie (Michigan State University)
- Alex Kent (PALS)
- Michael Bolam (University of Pittsburgh)
- Kelli Babcock (University of Toronto Libraries)
- Allison Brown (University of Otago Library)
- Kirsta Stapelfeldt (University of Toronto)
- Becky Yoose (Grinnell College Libraries)
- Ernie Gillis (Berklee College of Music Archives)
- Sofia Becerra-Licha (Berklee College of Music Archives)
- Gloria Gonzalez (UCLA)
- Stephen Davison (UCLA)
- Jane Monson (University of Northern Colorado)
- Andrew Berger (Computer History Museum)
- Caleb Derven (University of Limerick)
- Kilian Amrhein (Zuse Institute Berlin)
- Eric Luhrs (Lafayette College Libraries)
- James R. Griffin III (Lafayette College Libraries)
- Thomas Goodnow (Lafayette College Libraries)
- Alex Garnet (Simon Fraser University)
- Angela Dappert (Digital Preservation Coalition)
- Courtney Mumma (Artefactual)

What else are we thinking about?

Roadmap

Checksum checker

- check fits checksum
- check objects outside islandora - in bags?
- check objects upon ingest that has pre-existing checksums

Nick Ruest: policy and documentation <http://digital.library.yorku.ca/tags/digital-preservation-policy> <http://digital.library.yorku.ca/tags/preservation-action-plan> <https://spotdocs.scholarsportal.info/display/ODCC/Policies%2C+procedures%2C+and+best+practices+documents>

PREMIS

Becky Yoose: On D6 but will be implementing various preservation modules Will be putting OJS and OCS content into Islandora

*more rights metadata sources -- we just do dc.rights now derivative

Nathan Books: I7 and Fedora, DSpace, Developing formal preservation policy

FITS

- Format identification

Ernie Gillis and Sofia Becerra-Licha: No formal digital preservation policy in place -- archives formally established in October 2012 with hiring of first college archivist, so we're just getting started. New to Islandora/Fedora. Starting by using it to set up repository for master's capstone projects, which involve a variety of media formats. Media requirements will be heavy in audio and video content. Projects created with extensive multimedia (including audio, video, animation, etc) will be uploaded by students and curated professional staff.

BagIt

- Batch ingest/import from bags
- External referrals to archival objects in Bag
- Use manifest to check the checksum from the manifest with the
- Provide a way to feed "plugin" parameters into Drush or in some

Mark Jordan: Contentdm to Islandora migration by end of calendar year; will be using FITS and checksum functionality in Islandora for sure Will be migration our IR to Islandora after the Cdm migration, and this content will be pushed to Archivematica for AIP creation and mgnt

Ryan Gjerde and Sasha Griffin: Focusing on digital preservation with regards to records management and born digital records Also marrying this with need for asset management on campus

PDF Solution Pack

- Create PDF/A upon ingest

Andrew Berger: Just starting out with Islandora and digital preservation Highest priority getting control over what we have Lots of large (HD) digital video files (production quality oral history, exhibit, event video, PBS show). Also looking at Archivematica

Some other ideas...

FedoraApi.php PHP
Last indexed on Feb 21

```
642 * - formatURI: the format URI of the datastream.  
643 * - checksumType: the algorithm used to compute the checksum. One of  
...  
707 $this->connection->addParamArray($request, $separator, $params, 'formatURI');  
708 $this->connection->addParamArray($request, $separator, $params, 'checksumType');
```

Datastream.php PHP
Last indexed on Jun 12

```
1431 break;  
1432  
1433 case 'set':  
1434     $this->modifyDatastream(array('formatURI' => $value));  
1435     break;  
...  
1435 break;  
1436  
1437 case 'unset':  
1438     $this->modifyDatastream(array('formatURI' => ''));  
1439     break;
```

Object.php PHP
Last indexed on Oct 2, 2013

```
839 $dsState => $ds->state,  
840 'formatURI' => $ds->format,  
841 'checksumType' => $ds->checksumType,  
842 'mimeType' => $ds->mimeType,
```

tests/FedoraApiTest.php PHP
Last indexed on Nov 26, 2013

```
936 $this->apim->modifyDatastream($pid, $dsid, array('formatURI' => 'testtesttest'));  
937 $actual = $this->apim->getDatastream($pid, $dsid);
```

How are these search results? Tell us!

formatURI & Pronom?

info:pronom/fmt/7

info:pronom/fmt/8

info:pronom/fmt/9

info:pronom/fmt/10

```
- <fits xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version=  
- <identification>  
  - <identity format="Tagged Image File Format" mimetype="image/tiff" toolname="FITS" toolversion="0.6.3"/>  
    <tool toolname="Jhove" toolversion="1.5"/>  
    <tool toolname="file utility" toolversion="5.09"/>  
    <tool toolname="Exiftool" toolversion="9.13"/>  
    <tool toolname="Droid" toolversion="3.0"/>  
    <tool toolname="NLNZ Metadata Extractor" toolversion="3.4GA"/>  
    <tool toolname="ffident" toolversion="0.2"/>  
    <version toolname="Jhove" toolversion="1.5" status="CONFLICT">5.0</version>  
    <version toolname="Droid" toolversion="3.0" status="CONFLICT">3</version>  
    <version toolname="Droid" toolversion="3.0" status="CONFLICT">4</version>  
    <version toolname="Droid" toolversion="3.0" status="CONFLICT">6</version>  
    <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/7</externalIdentifier>  
    <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/8</externalIdentifier>  
    <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/9</externalIdentifier>  
    <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/10</externalIdentifier>  
  </identity>  
</identification>
```

Preservation Documentation

- [Generic, York University Libraries.](#)
- [Thank you Scholars Portal!](#)
- Develop a standard set of preservation policies, action plans and related documentation relevant to an Islandora install.
- Ideally a module that an Islandora Repository administrator could install and then customize the resulting documentation as needed.



Search input field with search button

Advanced search

Preservation Policy

- [Backup Plan](#)
- [Content Types](#)
- [Definition of AIP](#)
- [Definition of DIP](#)
- [Definition of SIP](#)
- [Designated Community Definition](#)
- [Digital Preservation Implementation Plan](#)
- [Digital Preservation Strategic Plan](#)
- [Environmental Monitoring of Preservation Formats](#)
- [Fixity procedures](#)
- [Registry of file formats](#)
- [Rights Policy](#)
- [URI Policy](#)

Preservation Action Plans

- [Audio](#)
- [Image](#)
- [Theses](#)
- [Video](#)
- [Web Archives](#)

Username *

Username input field

Password *

Password input field

Log in

Documentation

URI Policy

Submitted by nruest on Wed, 01/22/2014 - 19:06

York University Library URI Policy

Policy Statement

URIs created by York University Digital Library

- York University Digital Library uses a systematic convention to generate unambiguously unique identification for digital objects within its repository. This convention will create a stable name or reference to an object that can be permanently associated with that object, regardless of future changes to organizational structure or to digital access protocols.
- This is in conformance with section 4.2.4 of Metrics for Digital Repository Audit and Certification (CCSDS, June 2009) which states that a compliant repository "shall have and use a convention that generates persistent, unique identifiers for all AIPs" and "its components."
- This convention will ensure that "each AIP can be unambiguously found in the future" and that "each AIP can be distinguished from all other AIPs in the repository"

Implementation

Islandora object

York University Digital Library canonical URIs are consistently constructed in the following manner:

- `/islandora/object/PID`

These URIs are aliased using [Islandora Pathauto](#) to the following pattern:

- `[fedora:pid]/[fedora:label]`

Example:

- **Photograph:** New Woodbine : racehorses train for opening of season
- **Canonical URI:** <http://digital.library.yorku.ca/islandora/object/yul:88675>
- **Aliases URL:** <http://digital.library.yorku.ca/yul-88675/new-woodbine-racehorses-train-opening-season>

Islandora object datastream

York University Digital Library object datastream canonical URIs are consistently constructed in the following manner:

- `/islandora/object/PID/datastream/DATASTREAM_NAME/view`
- `/islandora/object/PID/datastream/DATASTREAM_NAME/download`
- `[fedora:pid]/[fedora:label]/datastream/DATASTREAM_NAME/view`
- `[fedora:pid]/[fedora:label]/datastream/DATASTREAM_NAME/download`

Example:

- **Photograph:** New Woodbine : racehorses train for opening of season

Preservation vs Exhibits

“There is no preservation without access”