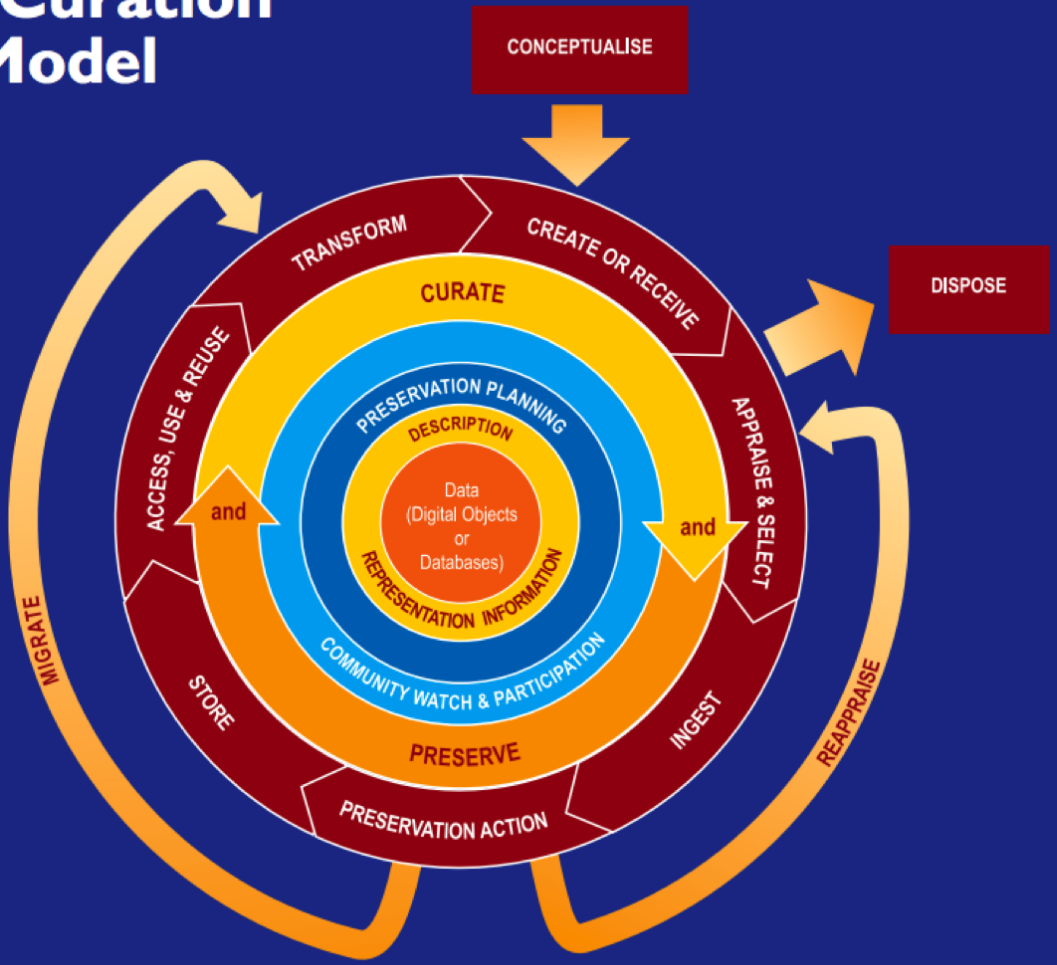
Preservation Policy for Humans

Really? Policy first thing in the morning?
...at least it isn’t right after lunch.


metro toronto
convention centre

The DCC Curation Lifecycle Model



Where should I start?

What do you care about?

Identify it!

...and maybe prioritize it too!

What can preserve?

Immediate, low hanging fruit

What could preserve?

Scope, rights

What can our infrastructure support?

Size, type

What could our infrastructure support?

Size, type

We've identified some objects.

What next?

Preservation actions

What’s that?

...anything we do to identify/maintain the

Integrity
Authenticity
Usability

Preservation Policy

Preservation Action Plans

Image - Preservation Action Plan

Submitted by nruest on Wed, 10/02/2013 - 11:47

Introduction

This document describes the preservation plan for image content in the York University Digital Library. Most of the image content content is from the Clara Thomas Archives and Special Collections. The preservation plan for image content follows from policies and practices described in the Digital Preservation Strategic Plan and the Digital Preservation Implementation Plan. This document explains practical steps that York University Libraries take to preserve the intellectual content of image in digital format. It outlines the basic tools, methods, and standards used for the long-term preservation of image content.

Content Formats

For the preservation of image content, York University Libraries require uncompressed TIF versions of the content, and descriptive metadata. During the ingest process, derivatives are created for display. York University Libraries continuously monitors developments in file formats to determine if and when formats require migration (see Environmental Monitoring of Preservation Formats).

SIP Format

Image SIPs (see Definition of SIP) generally consist of a TIF file, and an associated MODS descriptive metadata file.

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format and validation using FITS. The output of the FITS identification and validation processes are recorded to a techincal metadata datastream (TECHMD_FITS) that is associated with the object in the repository.

Content Excluded

York University Libraries do not ingest files that are not referenced (either as part of a representation or as associated datastreams) in the associated metadata. As the SIP is retained, these files can later be ingested if necessary.

Format Normalization

There is no format normailziation if the submitted object is a TIF.

Metadata Normalization

When necessary, York University Libraries crosswalk descriptive metadata from MODS to Dublin Core. The repository creates preservation metadata for each file. The preservation level, explained in the Digital Preservation Implementation Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

For the Full Preservation level for image, currently the acceptable formats are TIF. Image submissions may be JPG or PNG format, however they will be preserved at the Bit-level Preservation level.

Tags:
Documentation

Say what you do, do what you say!

Consistent, predictable… based on a plan.

Wait… how do you account for that?

Hint: It’s based on the plan part.

Planning!

In three simple steps...

1. **What is our general approach?**

2. **What tools do we have available?**

3. **How do we apply said tools?**

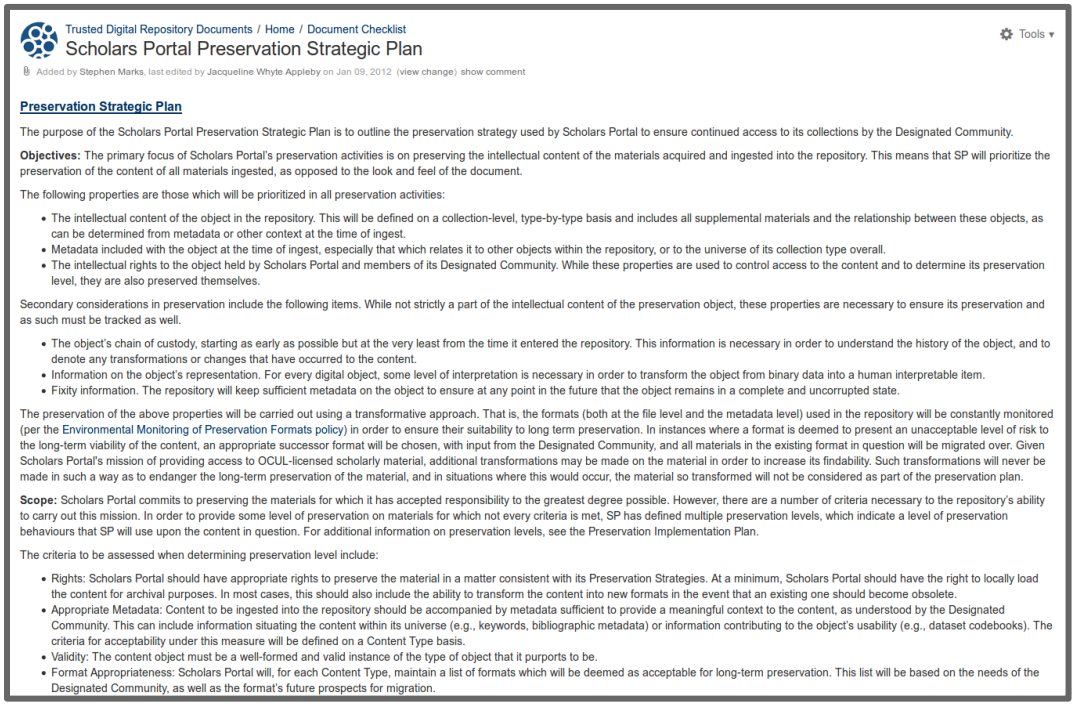
cough
Use cases!
cough

General Approach

Preservation Strategic Plan

Says things like e.g. “These are the important characteristics of the stuff we care about.”

Examples



Trusted Digital Repository Documents / Home / Document Checklist

Tools

Scholars Portal Preservation Strategic Plan

Added by **Stephen Marks**, last edited by **Jacqueline Whyte Appleby** on Jan 09, 2012 (view change) show comment

Preservation Strategic Plan

The purpose of the Scholars Portal Preservation Strategic Plan is to outline the preservation strategy used by Scholars Portal to ensure continued access to its collections by the Designated Community.

Objectives: The primary focus of Scholars Portal's preservation activities is on preserving the intellectual content of the materials acquired and ingested into the repository. This means that SP will prioritize the preservation of the content of all materials ingested, as opposed to the look and feel of the document.

The following properties are those which will be prioritized in all preservation activities:

- The intellectual content of the object in the repository. This will be defined on a collection-level, type-by-type basis and includes all supplemental materials and the relationship between these objects, as can be determined from metadata or other context at the time of ingest.
- Metadata included with the object at the time of ingest, especially that which relates it to other objects within the repository, or to the universe of its collection type overall.
- The intellectual rights to the object held by Scholars Portal and members of its Designated Community. While these properties are used to control access to the content and to determine its preservation level, they are also preserved themselves.

Secondary considerations in preservation include the following items. While not strictly a part of the intellectual content of the preservation object, these properties are necessary to ensure its preservation and as such must be tracked as well.

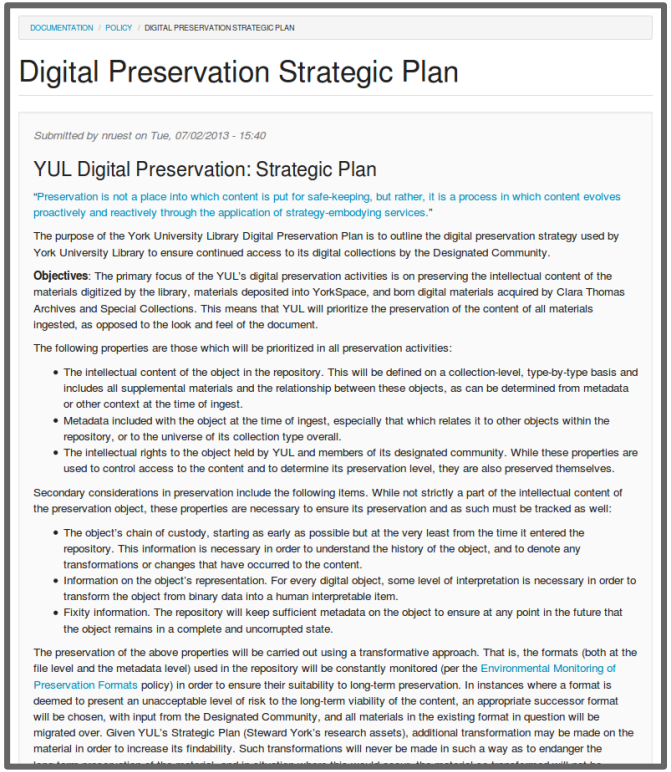
- The object's chain of custody, starting as early as possible but at the very least from the time it entered the repository. This information is necessary in order to understand the history of the object, and to denote any transformations or changes that have occurred to the content.
- Information on the object's representation. For every digital object, some level of interpretation is necessary in order to transform the object from binary data into a human interpretable item.
- Fixity information. The repository will keep sufficient metadata on the object to ensure at any point in the future that the object remains in a complete and uncorrupted state.

The preservation of the above properties will be carried out using a transformative approach. That is, the formats (both at the file level and the metadata level) used in the repository will be constantly monitored (per the Environmental Monitoring of Preservation Formats policy) in order to ensure their suitability to long term preservation. In instances where a format is deemed to present an unacceptable level of risk to the long-term viability of the content, an appropriate successor format will be chosen, with input from the Designated Community, and all materials in the existing format in question will be migrated over. Given Scholars Portal's mission of providing access to OCUL-licensed scholarly material, additional transformations may be made on the material in order to increase its findability. Such transformations will never be made in such a way as to endanger the long-term preservation of the material, and in situations where this would occur, the material so transformed will not be considered as part of the preservation plan.

Scope: Scholars Portal commits to preserving the materials for which it has accepted responsibility to the greatest degree possible. However, there are a number of criteria necessary to the repository's ability to carry out this mission. In order to provide some level of preservation on materials for which not every criteria is met, SP has defined multiple preservation levels, which indicate a level of preservation behaviours that SP will use upon the content in question. For additional information on preservation levels, see the Preservation Implementation Plan.

The criteria to be assessed when determining preservation level include:

- Rights: Scholars Portal should have appropriate rights to preserve the material in a matter consistent with its Preservation Strategies. At a minimum, Scholars Portal should have the right to locally load the content for archival purposes. In most cases, this should also include the ability to transform the content into new formats in the event that an existing one should become obsolete.
- Appropriate Metadata: Content to be ingested into the repository should be accompanied by metadata sufficient to provide a meaningful context to the content, as understood by the Designated Community. This can include information situating the content within its universe (e.g., keywords, bibliographic metadata) or information contributing to the object's usability (e.g., dataset codebooks). The criteria for acceptability under this measure will be defined on a Content Type basis.
- Validity: The content object must be a well-formed and valid instance of the type of object that it purports to be.
- Format Appropriateness: Scholars Portal will, for each Content Type, maintain a list of formats which will be deemed as acceptable for long-term preservation. This list will be based on the needs of the Designated Community, as well as the format's future prospects for migration.



DOCUMENTATION / POLICY / DIGITAL PRESERVATION STRATEGIC PLAN

Digital Preservation Strategic Plan

Submitted by nruest on Tue, 07/02/2013 - 15:40

YUL Digital Preservation: Strategic Plan

"Preservation is not a place into which content is put for safe-keeping, but rather, it is a process in which content evolves proactively and reactively through the application of strategy-embodying services."

The purpose of the York University Library Digital Preservation Plan is to outline the digital preservation strategy used by York University Library to ensure continued access to its digital collections by the Designated Community.

Objectives: The primary focus of the YUL's digital preservation activities is on preserving the intellectual content of the materials digitized by the library, materials deposited into YorkSpace, and born digital materials acquired by Clara Thomas Archives and Special Collections. This means that YUL will prioritize the preservation of the content of all materials ingested, as opposed to the look and feel of the document.

The following properties are those which will be prioritized in all preservation activities:

- The intellectual content of the object in the repository. This will be defined on a collection-level, type-by-type basis and includes all supplemental materials and the relationship between these objects, as can be determined from metadata or other context at the time of ingest.
- Metadata included with the object at the time of ingest, especially that which relates it to other objects within the repository, or to the universe of its collection type overall.
- The intellectual rights to the object held by YUL and members of its designated community. While these properties are used to control access to the content and to determine its preservation level, they are also preserved themselves.

Secondary considerations in preservation include the following items. While not strictly a part of the intellectual content of the preservation object, these properties are necessary to ensure its preservation and as such must be tracked as well:

- The object's chain of custody, starting as early as possible but at the very least from the time it entered the repository. This information is necessary in order to understand the history of the object, and to denote any transformations or changes that have occurred to the content.
- Information on the object's representation. For every digital object, some level of interpretation is necessary in order to transform the object from binary data into a human interpretable item.
- Fixity information. The repository will keep sufficient metadata on the object to ensure at any point in the future that the object remains in a complete and uncorrupted state.

The preservation of the above properties will be carried out using a transformative approach. That is, the formats (both at the file level and the metadata level) used in the repository will be constantly monitored (per the Environmental Monitoring of Preservation Formats policy) in order to ensure their suitability to long-term preservation. In instances where a format is deemed to present an unacceptable level of risk to the long-term viability of the content, an appropriate successor format will be chosen, with input from the Designated Community, and all materials in the existing format in question will be migrated over. Given YUL's Strategic Plan (Steward York's research assets), additional transformation may be made on the material in order to increase its findability. Such transformations will never be made in such a way as to endanger the

Tools in the Toolbox

Given our overall goals, what tools do we have that can support it?

Example Tools

Archival Formats

Content Formats

For the preservation of audio content, York University Libraries require WAV or FLAC versions of the content, and descriptive metadata. During the ingest process, derivatives are created for streaming. York University Libraries continuously monitors developments in file formats to determine if and when formats require migration (see Environmental Monitoring of Preservation Formats).

Integrity monitoring

Fixity procedures

Submitted by nruest on Thu, 06/27/2013 - 12:51

Policy Statement

York University Library are committed to maintaining the integrity of objects in its care. This includes creating checksums for all archival format objects -- plus associated datastreams -- ingested into the repository, and regular fixity checking of those objects.

Implementation

At the time of ingest an SHA1 checksum value is calculated for the archival format object, and is stored along the object in the repository.

Daily, a set number of files in the repository will have their current checksum calculated (using a single checksum) and compared to this stored value, which is expected to match. In cases where the calculated and stored values do not match, this is reported to the repository manager.

Tags:
Documentation
Digital Preservation Policy

Format normalization

Format Normalization

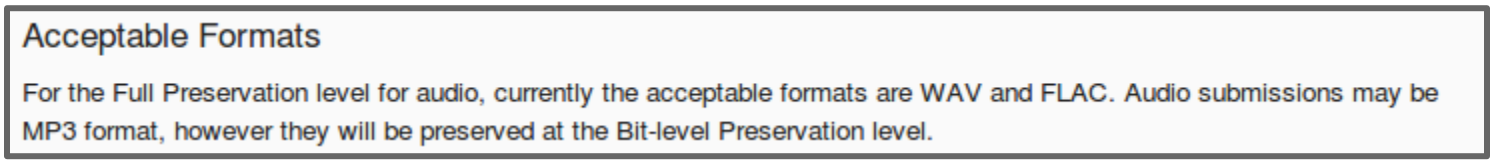
There is no format normailziation if the submitted object is a WAV or FLAC.

Format Normalization

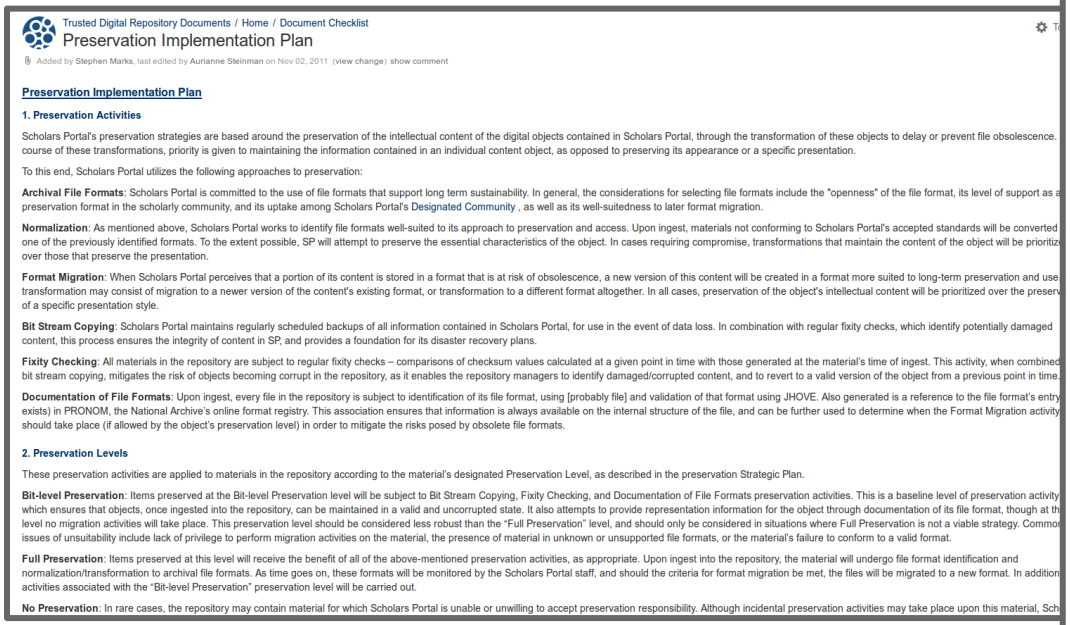
Upon ingest, the publisher's XML/SGML is converted to a Scholars Portal version of NLM XML. Where possible and when desirable, files that do not conform to Scholars Portal's preferred formats will be converted to preferred formats.

Other considerations

Are there different levels of preservation we can offer?



Preservation Implementation Plan



Trusted Digital Repository Documents / Home / Document Checklist

Preservation Implementation Plan

Added by Stephen Marks, last edited by Aurianne Steinman on Nov 02, 2011 (view change) show comment

Preservation Implementation Plan

1. Preservation Activities

Scholars Portal's preservation strategies are based around the preservation of the intellectual content of the digital objects contained in Scholars Portal, through the transformation of these objects to delay or prevent file obsolescence. course of these transformations, priority is given to maintaining the information contained in an individual content object, as opposed to preserving its appearance or a specific presentation.

To this end, Scholars Portal utilizes the following approaches to preservation:

Archival File Formats: Scholars Portal is committed to the use of file formats that support long term sustainability. In general, the considerations for selecting file formats include the "openness" of the file format, its level of support as a preservation format in the scholarly community, and its uptake among Scholars Portal's Designated Community , as well as its well-suitedness to later format migration.

Normalization: As mentioned above, Scholars Portal works to identify file formats well-suited to its approach to preservation and access. Upon ingest, materials not conforming to Scholars Portal's accepted standards will be converted one of the previously identified formats. To the extent possible, SP will attempt to preserve the essential characteristics of the object. In cases requiring compromise, transformations that maintain the content of the object will be prioritiz over those that preserve the presentation.

Format Migration: When Scholars Portal perceives that a portion of its content is stored in a format that is at risk of obsolescence, a new version of this content will be created in a format more suited to long-term preservation and use transformation may consist of migration to a newer version of the content's existing format, or transformation to a different format altogether. In all cases, preservation of the object's intellectual content will be prioritized over the preserv of a specific presentation style.

Bit Stream Copying: Scholars Portal maintains regularly scheduled backups of all information contained in Scholars Portal, for use in the event of data loss. In combination with regular fixity checks, which identify potentially damaged content, this process ensures the integrity of content in SP, and provides a foundation for its disaster recovery plans.

Fixity Checking: All materials in the repository are subject to regular fixity checks – comparisons of checksum values calculated at a given point in time with those generated at the material's time of ingest. This activity, when combined bit stream copying, mitigates the risk of objects becoming corrupt in the repository, as it enables the repository managers to identify damaged/corrupted content, and to revert to a valid version of the object from a previous point in time.

Documentation of File Formats: Upon ingest, every file in the repository is subject to identification of its file format, using [probably file] and validation of that format using JHOVE. Also generated is a reference to the file format's entry exists) in PRONOM, the National Archive's online format registry. This association ensures that information is always available on the internal structure of the file, and can be further used to determine when the Format Migration activity should take place (if allowed by the object's preservation level) in order to mitigate the risks posed by obsolete file formats.

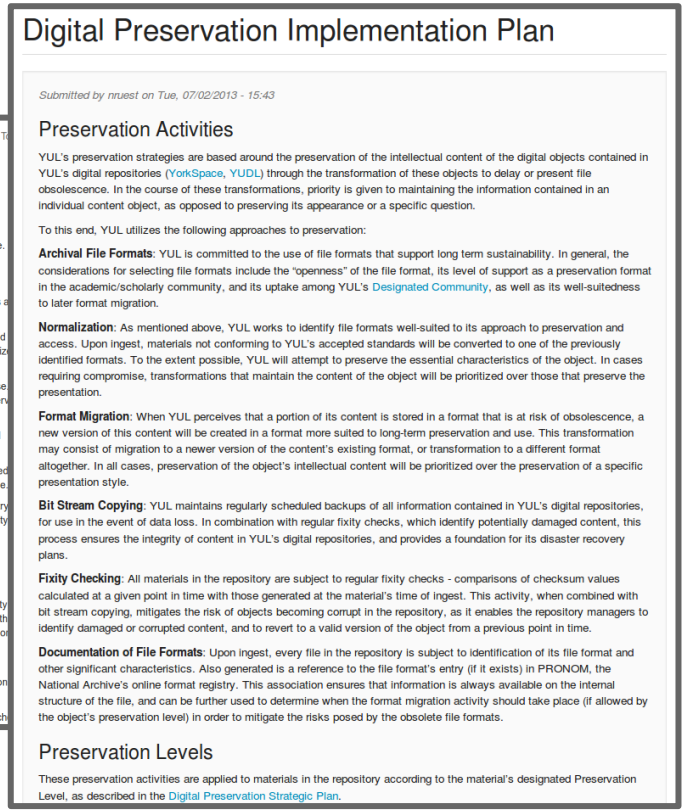
2. Preservation Levels

These preservation activities are applied to materials in the repository according to the material's designated Preservation Level, as described in the preservation Strategic Plan.

Bit-level Preservation: Items preserved at the Bit-level Preservation level will be subject to Bit Stream Copying, Fixity Checking, and Documentation of File Formats preservation activities. This is a baseline level of preservation activity which ensures that objects, once ingested into the repository, can be maintained in a valid and uncorrupted state. It also attempts to provide representation information for the object through documentation of its file format, though at th level no migration activities will take place. This preservation level should be considered less robust than the "Full Preservation" level, and should only be considered in situations where Full Preservation is not a viable strategy. Commo issues of unsuitability include lack of privilege to perform migration activities on the material, the presence of material in unknown or unsupported file formats, or the material's failure to conform to a valid format.

Full Preservation: Items preserved at this level will receive the benefit of all of the above-mentioned preservation activities, as appropriate. Upon ingest into the repository, the material will undergo file format identification and normalization/transformation to archival file formats. As time goes on, these formats will be monitored by the Scholars Portal staff, and should the criteria for format migration be met, the files will be migrated to a new format. In addition activities associated with the "Bit-level Preservation" preservation level will be carried out.

No Preservation: In rare cases, the repository may contain material for which Scholars Portal is unable or unwilling to accept preservation responsibility. Although incidental preservation activities may take place upon this material, Sch



Digital Preservation Implementation Plan

Submitted by nruest on Tue, 07/02/2013 - 15:43

Preservation Activities

YUL's preservation strategies are based around the preservation of the intellectual content of the digital objects contained in YUL's digital repositories (YorkSpace, YUDL) through the transformation of these objects to delay or present file obsolescence. In the course of these transformations, priority is given to maintaining the information contained in an individual content object, as opposed to preserving its appearance or a specific question.

To this end, YUL utilizes the following approaches to preservation:

Archival File Formats: YUL is committed to the use of file formats that support long term sustainability. In general, the considerations for selecting file formats include the "openness" of the file format, its level of support as a preservation format in the academic/scholarly community, and its uptake among YUL's Designated Community, as well as its well-suitedness to later format migration.

Normalization: As mentioned above, YUL works to identify file formats well-suited to its approach to preservation and access. Upon ingest, materials not conforming to YUL's accepted standards will be converted to one of the previously identified formats. To the extent possible, YUL will attempt to preserve the essential characteristics of the object. In cases requiring compromise, transformations that maintain the content of the object will be prioritized over those that preserve the presentation.

Format Migration: When YUL perceives that a portion of its content is stored in a format that is at risk of obsolescence, a new version of this content will be created in a format more suited to long-term preservation and use. This transformation may consist of migration to a newer version of the content's existing format, or transformation to a different format altogether. In all cases, preservation of the object's intellectual content will be prioritized over the preservation of a specific presentation style.

Bit Stream Copying: YUL maintains regularly scheduled backups of all information contained in YUL's digital repositories, for use in the event of data loss. In combination with regular fixity checks, which identify potentially damaged content, this process ensures the integrity of content in YUL's digital repositories, and provides a foundation for its disaster recovery plans.

Fixity Checking: All materials in the repository are subject to regular fixity checks - comparisons of checksum values calculated at a given point in time with those generated at the material's time of ingest. This activity, when combined with bit stream copying, mitigates the risk of objects becoming corrupt in the repository, as it enables the repository managers to identify damaged or corrupted content, and to revert to a valid version of the object from a previous point in time.

Documentation of File Formats: Upon ingest, every file in the repository is subject to identification of its file format and other significant characteristics. Also generated is a reference to the file format's entry (if it exists) in PRONOM, the National Archive's online format registry. This association ensures that information is always available on the internal structure of the file, and can be further used to determine when the format migration activity should take place (if allowed by the object's preservation level) in order to mitigate the risks posed by the obsolete file formats.

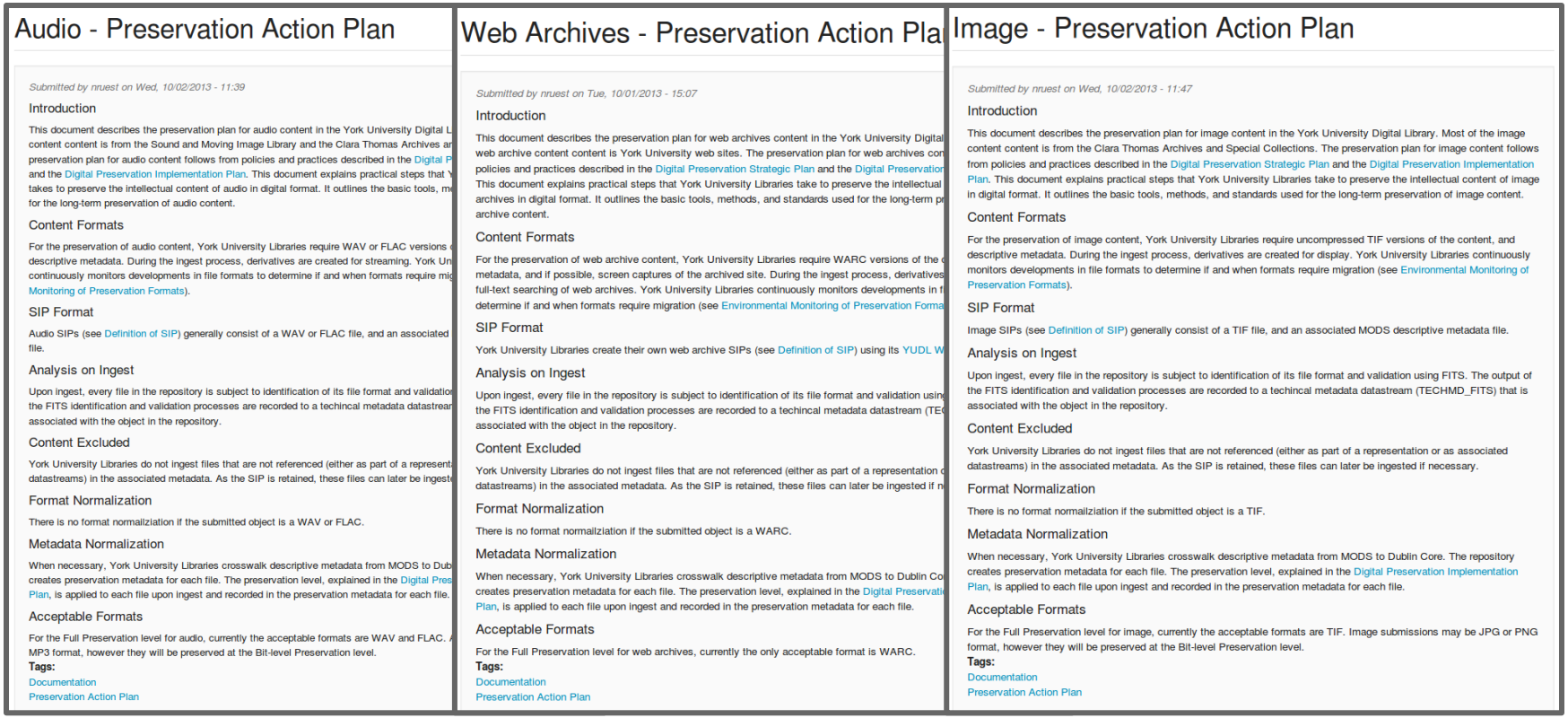
Preservation Levels

These preservation activities are applied to materials in the repository according to the material's designated Preservation Level, as described in the Digital Preservation Strategic Plan.

Given a particular use case….

How (specifically!) do we use the tools in our toolbox to enact our general approach on *this specific stuff*?

Preservation Action Plan examples



Audio - Preservation Action Plan

Submitted by nruest on Wed, 10/02/2013 - 11:39

Introduction

This document describes the preservation plan for audio content in the York University Digital L
content content is from the Sound and Moving Image Library and the Clara Thomas Archives ar
preservation plan for audio content follows from policies and practices described in the Digital P
and the Digital Preservation Implementation Plan. This document explains practical steps that Y
takes to preserve the intellectual content of audio in digital format. It outlines the basic tools, m
for the long-term preservation of audio content.

Content Formats

For the preservation of audio content, York University Libraries require WAV or FLAC versions
descriptive metadata. During the ingest process, derivatives are created for streaming. York Un
continuously monitors developments in file formats to determine if and when formats require mig
Monitoring of Preservation Formats).

SIP Format

Audio SIPs (see Definition of SIP) generally consist of a WAV or FLAC file, and an associated
file.

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format and validatio
the FITS identification and validation processes are recorded to a techincal metadata datastrean
associated with the object in the repository.

Content Excluded

York University Libraries do not ingest files that are not referenced (either as part of a represent
datastreams) in the associated metadata. As the SIP is retained, these files can later be ingest

Format Normalization

There is no format normailziation if the submitted object is a WAV or FLAC.

Metadata Normalization

When necessary, York University Libraries crosswalk descriptive metadata from MODS to Dub
creates preservation metadata for each file. The preservation level, explained in the Digital Pres
Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

For the Full Preservation level for audio, currently the acceptable formats are WAV and FLAC. A
MP3 format, however they will be preserved at the Bit-level Preservation level.

Tags:

Documentation
Preservation Action Plan

Web Archives - Preservation Action Pla

Submitted by nruest on Tue, 10/01/2013 - 15:07

Introduction

This document describes the preservation plan for web archives content in the York University Digital
web archive content content is York University web sites. The preservation plan for web archives con
policies and practices described in the Digital Preservation Strategic Plan and the Digital Preservation
This document explains practical steps that York University Libraries take to preserve the intellectual
archives in digital format. It outlines the basic tools, methods, and standards used for the long-term pr
archive content.

Content Formats

For the preservation of web archive content, York University Libraries require WARC versions of the
metadata, and if possible, screen captures of the archived site. During the ingest process, derivatives
full-text searching of web archives. York University Libraries continuously monitors developments in f
determine if and when formats require migration (see Environmental Monitoring of Preservation Forma

SIP Format

York University Libraries create their own web archive SIPs (see Definition of SIP) using its YUDL W

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format and validation using
the FITS identification and validation processes are recorded to a techincal metadata datastream (TE
associated with the object in the repository.

Content Excluded

York University Libraries do not ingest files that are not referenced (either as part of a representation
datastreams) in the associated metadata. As the SIP is retained, these files can later be ingested if n

Format Normalization

There is no format normailziation if the submitted object is a WARC.

Metadata Normalization

When necessary, York University Libraries crosswalk descriptive metadata from MODS to Dublin Co
creates preservation metadata for each file. The preservation level, explained in the Digital Preservati
Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

For the Full Preservation level for web archives, currently the only acceptable format is WARC.

Tags:

Documentation
Preservation Action Plan

Image - Preservation Action Plan

Submitted by nruest on Wed, 10/02/2013 - 11:47

Introduction

This document describes the preservation plan for image content in the York University Digital Library. Most of the image content content is from the Clara Thomas Archives and Special Collections. The preservation plan for image content follows from policies and practices described in the Digital Preservation Strategic Plan and the Digital Preservation Implementation Plan. This document explains practical steps that York University Libraries take to preserve the intellectual content of image in digital format. It outlines the basic tools, methods, and standards used for the long-term preservation of image content.

Content Formats

For the preservation of image content, York University Libraries require uncompressed TIF versions of the content, and descriptive metadata. During the ingest process, derivatives are created for display. York University Libraries continuously monitors developments in file formats to determine if and when formats require migration (see Environmental Monitoring of Preservation Formats).

SIP Format

Image SIPs (see Definition of SIP) generally consist of a TIF file, and an associated MODS descriptive metadata file.

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format and validation using FITS. The output of the FITS identification and validation processes are recorded to a techincal metadata datastream (TECHMD_FITS) that is associated with the object in the repository.

Content Excluded

York University Libraries do not ingest files that are not referenced (either as part of a representation or as associated datastreams) in the associated metadata. As the SIP is retained, these files can later be ingested if necessary.

Format Normalization

There is no format normailziation if the submitted object is a TIF.

Metadata Normalization

When necessary, York University Libraries crosswalk descriptive metadata from MODS to Dublin Core. The repository creates preservation metadata for each file. The preservation level, explained in the Digital Preservation Implementation Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

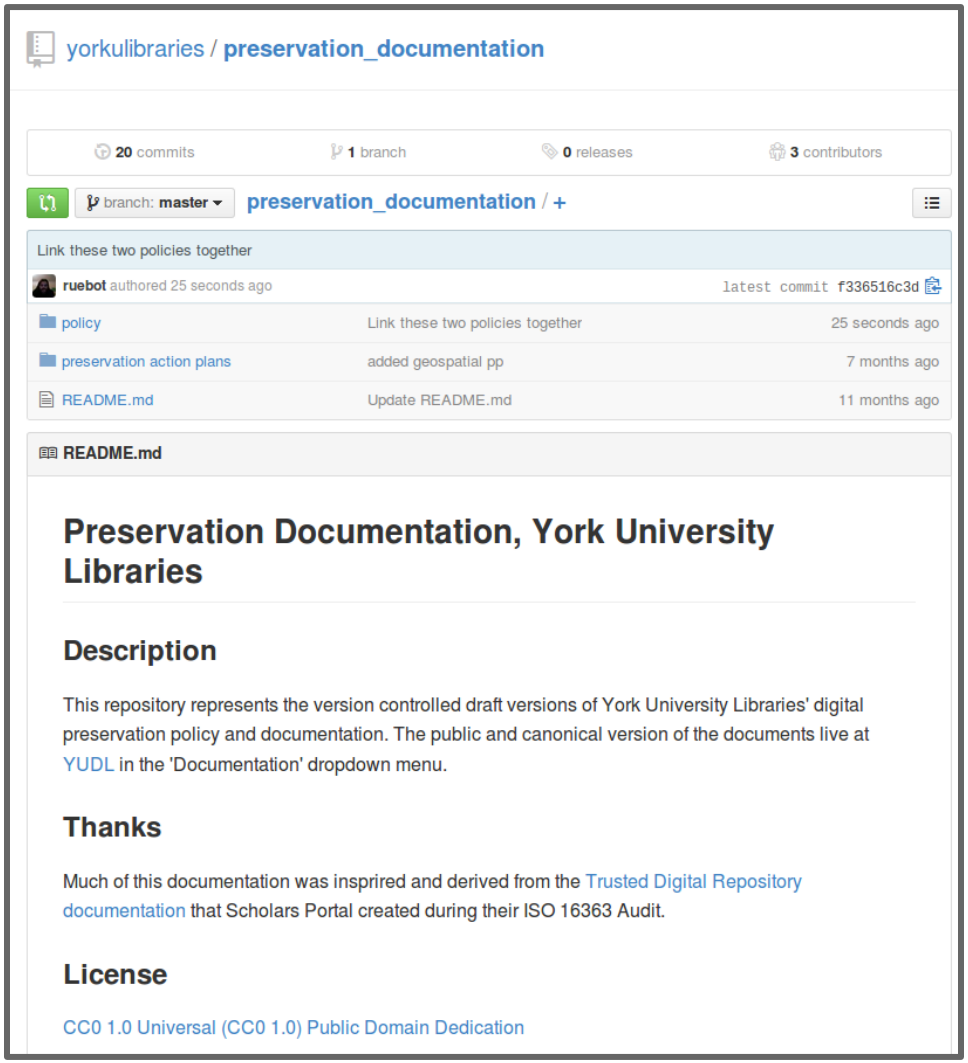
For the Full Preservation level for image, currently the acceptable formats are TIF. Image submissions may be JPG or PNG format, however they will be preserved at the Bit-level Preservation level.

Tags:

Documentation
Preservation Action Plan

Parting thoughts...

RELEASE EARLY
RELEASE OFTEN



yorkulibraries / preservation_documentation

20 commits | 1 branch | 0 releases | 3 contributors

branch: master ▾ preservation_documentation / +

Link these two policies together

ruebot authored 25 seconds ago — latest commit f336516c3d

policy	Link these two policies together	25 seconds ago
preservation action plans	added geospatial pp	7 months ago
README.md	Update README.md	11 months ago

README.md

Preservation Documentation, York University Libraries

Description

This repository represents the version controlled draft versions of York University Libraries' digital preservation policy and documentation. The public and canonical version of the documents live at YUDL in the 'Documentation' dropdown menu.

Thanks

Much of this documentation was insprired and derived from the Trusted Digital Repository documentation that Scholars Portal created during their ISO 16363 Audit.

License

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

Discussion, Questions?