The Lifecycle of a Digital Object

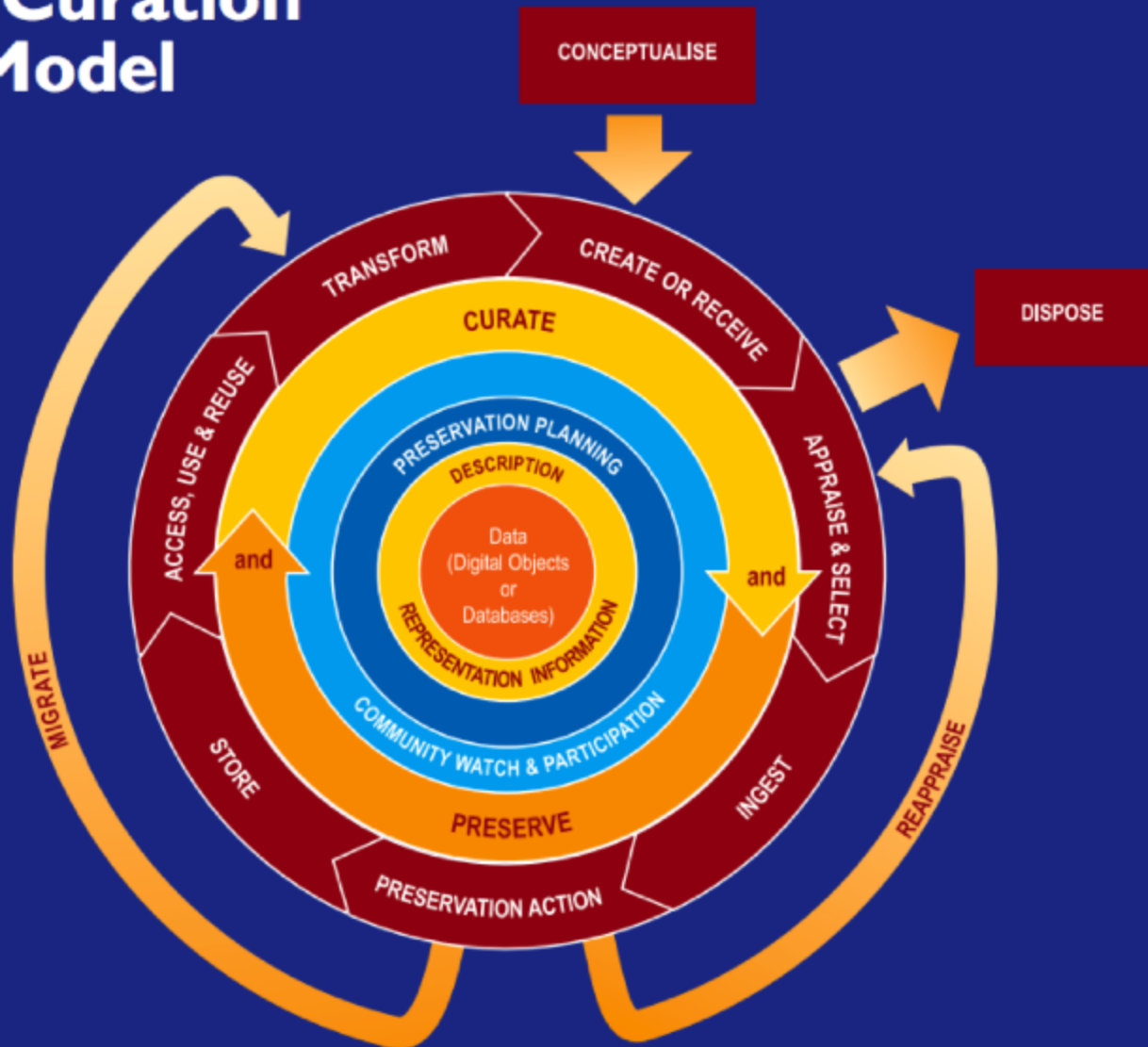## Digital Preservation 1.0

Nick Ruest & Steve Marks
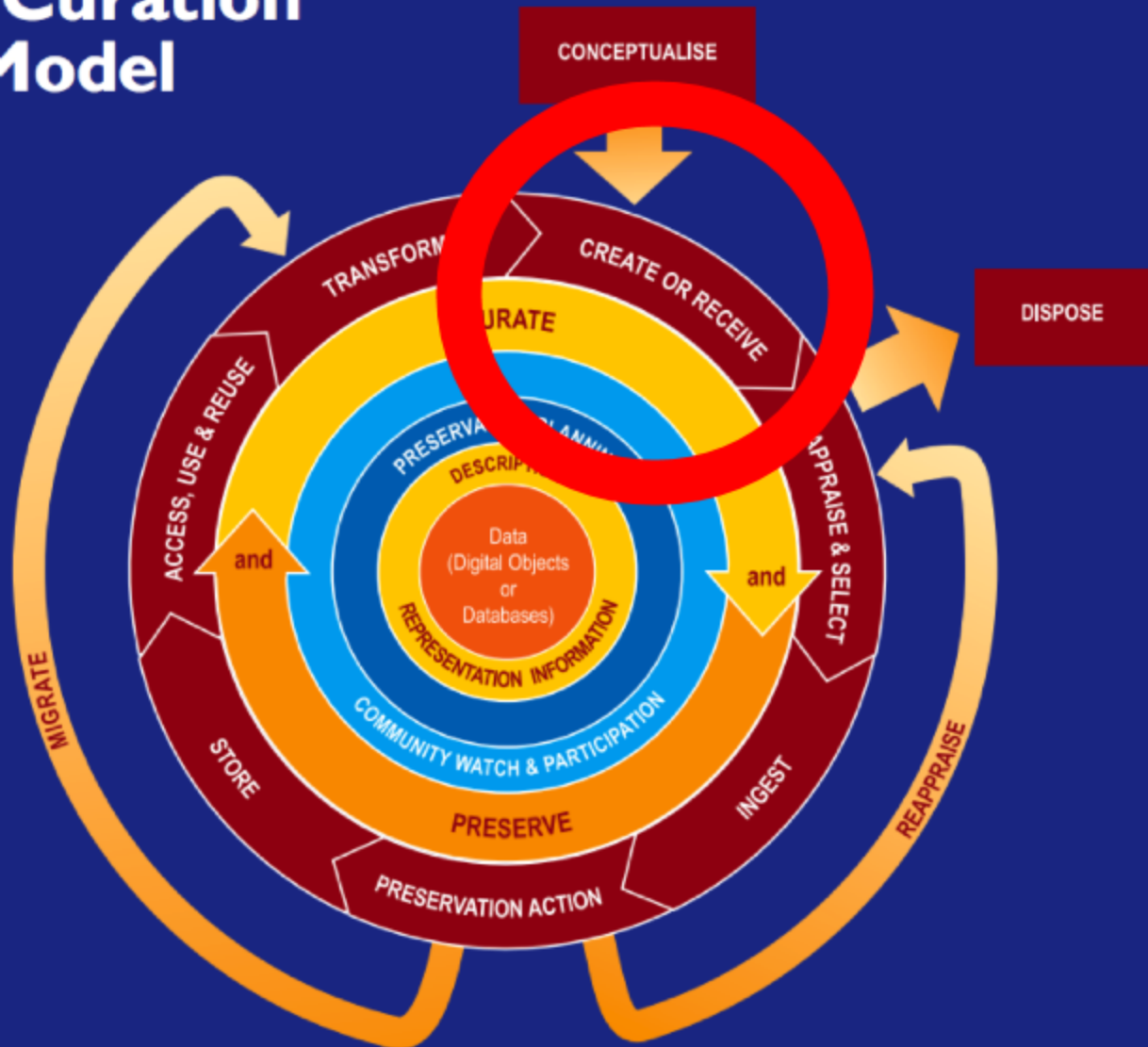August 13, 2014
educationinstitute.ca

What is Digital Curation?

"Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle."

Digital Curation Centre. "What is digital curation?", http://www.dcc.ac.uk/digital-curation/what-digital-curation

The DCC Curation Lifecycle Model

The DCC Curation Lifecycle Model

Create or Receive

How do we get well formed data?

We create *curation* ready data!

Develop

Influence

Create

Collect

Develop

POLICY!

such SIP specification

f4af8b5789576c000ce9105b25609bd6

Definition of SIP

Submitted by nruest on Tue, 07/02/2013 - 15:54

Submission Information Package (SIP)

- The information package that is delivered to York University Digital Library for use in the construction of one or more AIPs.
- The format of the SIP may vary from submitter to submitter, based on the submitters willingness and ability to provide the content and metadata in a specific format.
- For a given Content Type, any requirements or restrictions on the type of content that can be contained in the SIP will be described in that Content Type's Preservation Action Plan.

Tags:

Documentation

Digital Preservation Policy

Well formed data/objects?

What does that look like!?
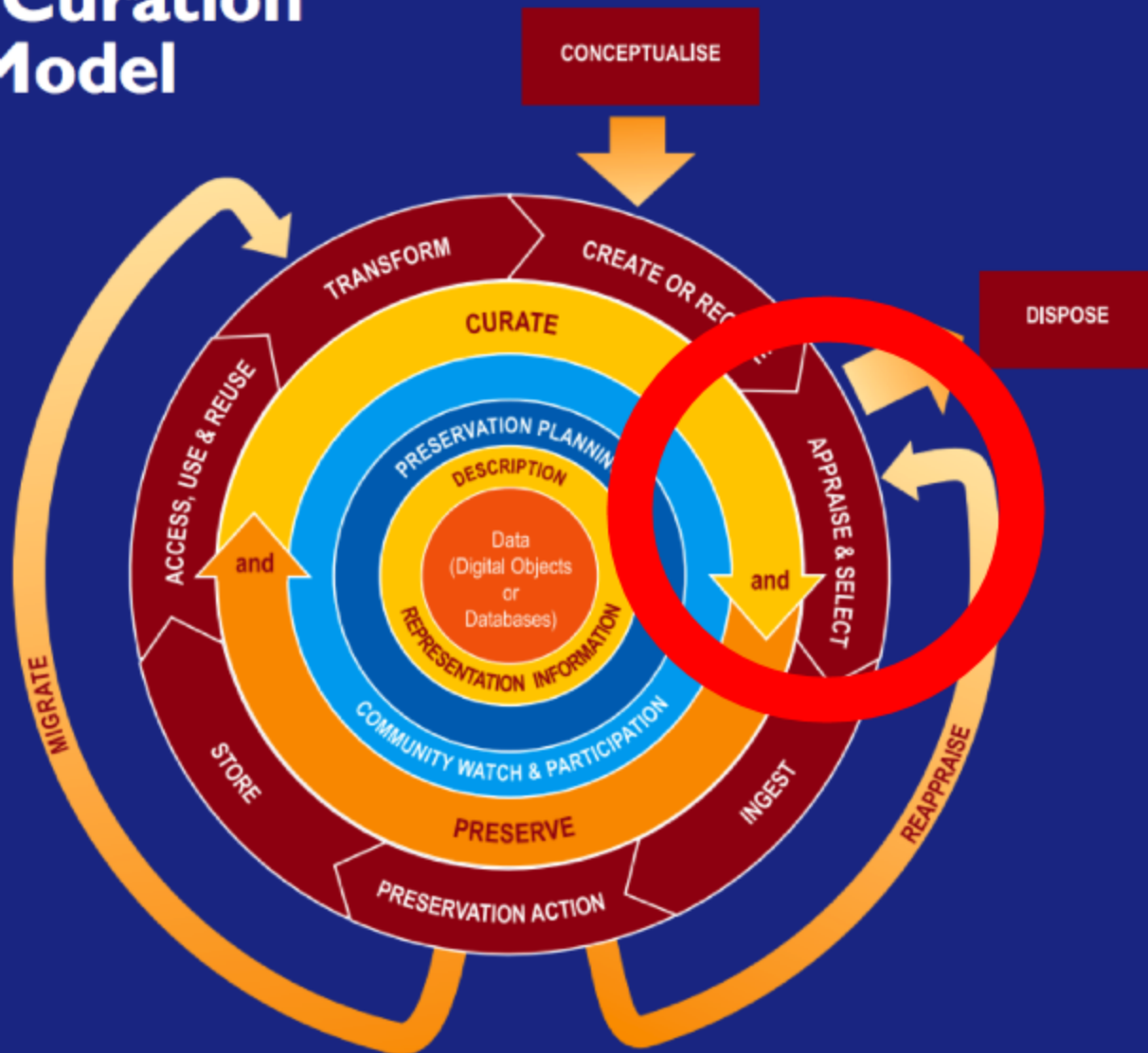
0000000 643c 7669 6320 616c 7373 223d 6f62 6472

Openness

Portability

Quality

Sample file format policy

The DCC Curation Lifecycle Model

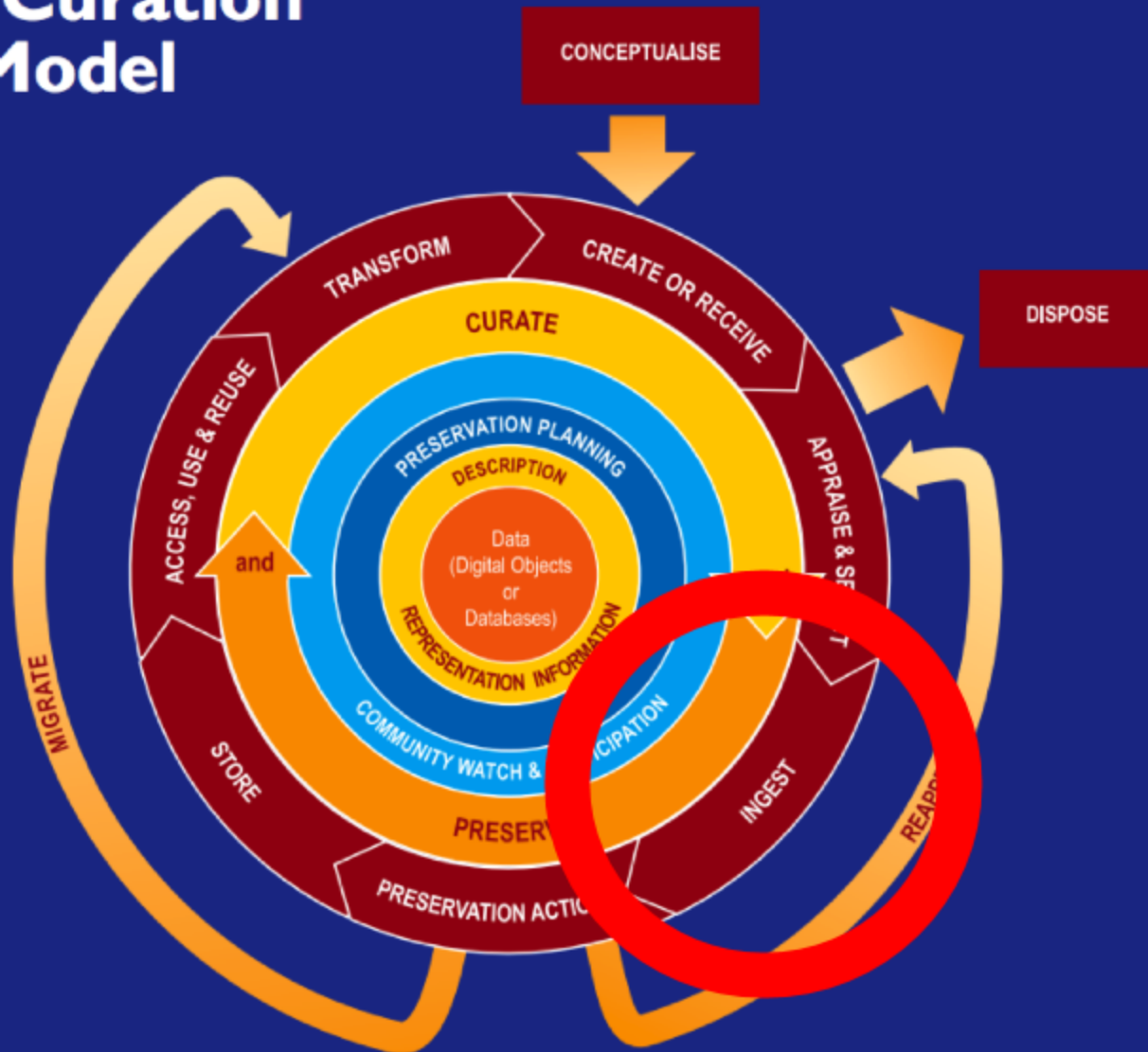Everything that we choose to preserve means that something else won't be.

How do we decide?

- Needs of users (Designated Community)

- Feasibility of preservation

- Legal and IP rights

- Criticality of data

- Presence of associated data/metadata

POLICY!

f4af8b5789576c000ce9105b25609bd6

The DCC Curation Lifecycle Model

SIPs to AIPs

Step 1: Submit

Step 2: Ingest

What's an AIP?

Reference Information

Provenance

Context Information

Fixity Information

Process!

1. Receive and accept SIP

2. Prepare the SIP for storage and management

3. Perform quality assurance activities on the SIP

4. Initiate format conversion to create the AIP

5. Generate AIP

POLICY!

Definition of AIP

f4af8b5789576c000ce9105b25609bd6

Definition of AIP

Submitted by nruest on Tue, 07/02/2013 - 15:54

Archival Information Package (AIP)

- The information package consisting of the Content Information (CI), Preservation Description Information (PDI), Packaging Information (PI), and Descriptive Information (DI) that is archived at York University Library.
- The level of content in a York University Digital Library AIP can vary, depending on the amount of content provided by the submitter.
- This description will use the OAIS Information Model to illustrate completeness of our conceptual model, and will describe, in general terms, what a York University Library AIP looks like.

Content Information (CI)

- The Content Data Object is generally stored with from the primary preservation metadata file, which is held in Fedora Commons.
- Representation Information is maintained, and contains information on the CDO's file format, version, and a reference to a format registry in order to provide information on how to interpret the file. See: registry of file formats

Preservation Description Information (PDI)

- *Reference Information* - Identifiers are stored for each object identifying it globally (e.g. YUL PID) and locally (e.g. URI).
- *Provenance Information* - Provenance metadata is maintained for each object that provides a history of preservation events in the object's lifetime, beginning at ingest into the YUL Digital Library repository and referencing any preservation activities taken on the object (e.g., replacement due to corruption, format migration, etc.).
- *Context Information* - As appropriate, information on how a CDO relates to other CDOs or to other conceptual entities. Examples of these relationships can include: a newer version of an object that supersedes an older one.
- *Fixity Information* - Fixity information is generated at the time of ingest in order to later determine whether or not the item remains in the same state as when it was ingested. This information can be used to determine integrity of an object being copied within the system (as in the case of a change in storage location), or for periodic integrity checks.

Packaging Information (PI)

- YUL preservation metadata packages both the descriptive and preservation metadata together.

Descriptive Information (DI)

- Depending on the type of CDO, the format of this descriptive metadata can vary (MODS or Dublin Core), but is selected to maximize findability. In all cases, the descriptive metadata will be recreated within the preservation metadata.

Tags:
Documentation
Digital Preservation Policy

Policy. Policy. Policy.

Sort & Identify

Acknowledge receipt

List data/objects
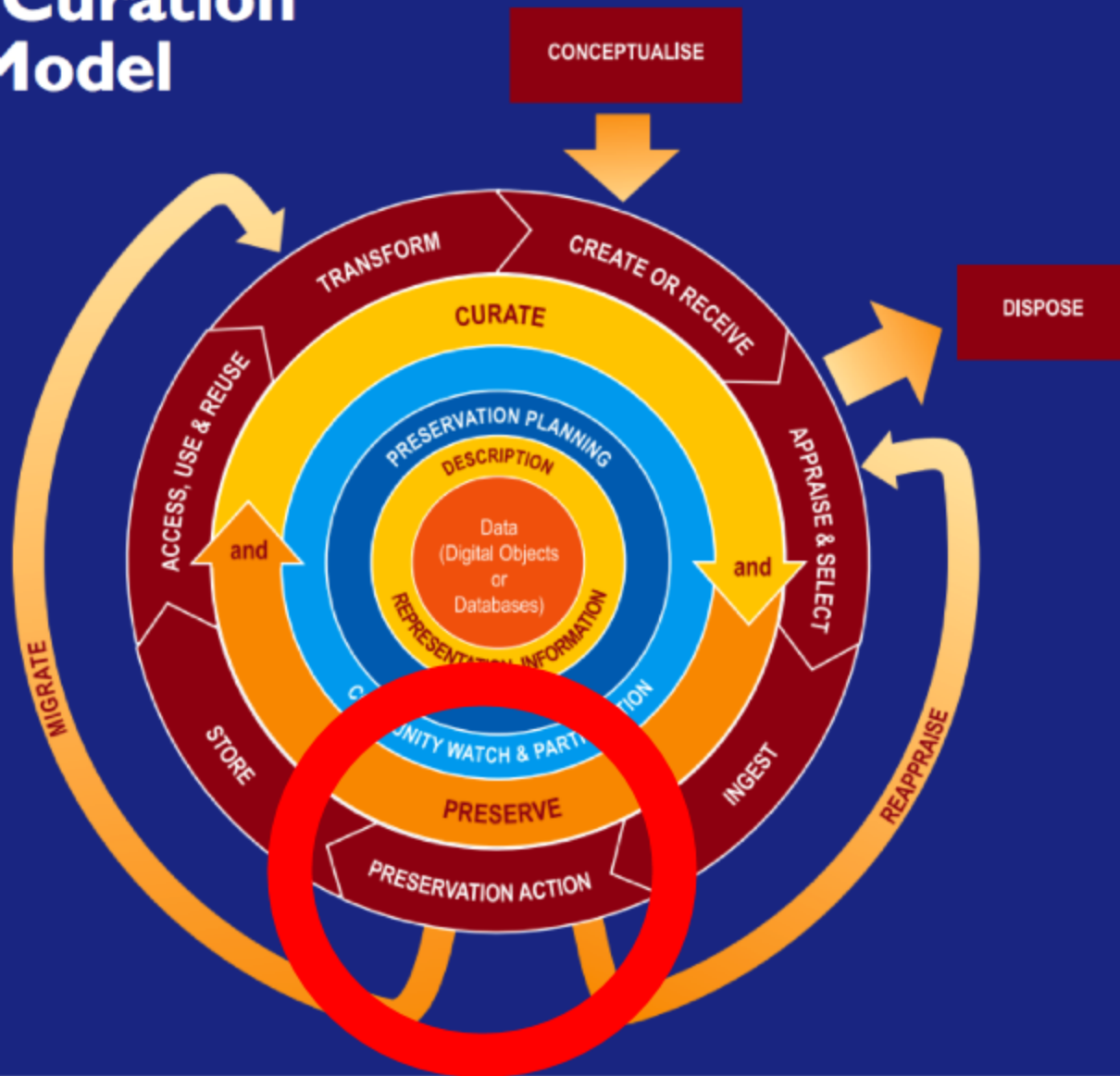
Depositor agreement

Uncompress

Decrypt

Virus & malware

Fixity

Permanent identifiers

Transform/derivatives

The DCC Curation Lifecycle Model

Preservation Action

Anything that we do to maintain the

- Integrity

- Authenticity

- Usability

of our content.

Usability

3 main strategies:

- Technology preservation

- Technology emulation

- Information migration

Case 1:

Disabled or young? Relative age and special education diagnoses in schools☆

Elizabeth Dhuey [a,*], Stephen Lipscomb [b,1]

[a] Centre for Industrial Relations and Human Resources, Department of Management, University of Toronto, 121 St. George Street, Toronto, Ontario, M5S 2E8 Canada
[b] Mathematica Policy Research, 955 Massachusetts Avenue, Suite 801, Cambridge, MA 02139, USA

ARTICLE INFO

ABSTRACT

This study extends recent findings of a relationship between the relative age of students among their peers and their probability of disability classification. Using three nationally representative surveys spanning 1988–2004 and grades K-10, we find that an additional month of relative age decreases the likelihood of receiving special education services by 2–5 percent. Relative age effects are strong for learning disabilities but not for other disabilities. We measure them for boys starting in kindergarten but not for girls until 3rd grade. We also measure them for white and Hispanic students but not for black students or differentially by socioeconomic quartiles. Results are consistent with the interpretation that disability assessments do not screen for the possibility that relatively young students are over-referred for evaluation. Lastly, we present suggestive evidence that math achievement gains due to disability classification may differentially benefit relatively young students.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Students with disabilities represented about 13.7 percent of the public school enrollment in the United States by 2005–2006, with about half diagnosed with learning disabilities.[2] All students with disabilities are entitled by law to a free and appropriate public education, which can be considerably more costly than educating students not classified with special needs. Spending on students with disabilities has been estimated to be 90 percent higher than for other students, on average (Chambers, Parrish, and Harr, 2004).[3] Special education spending also has grown faster than regular education spending since the 1980s, representing a larger share of district budgets (Lankford & Wyckoff, 1995; Parrish, 2001).

A recent study by Elder and Lubotsky (2009) finds compelling evidence that school officials may use relative standards in classifying children as having a disability. Their results indicate that children who start school at older biological ages are less likely to be classified with Attention Deficit Disorder (ADD) or Attention Deficit Hyperactivity Disorder (ADHD) by fifth grade.[4] The effects are large; starting school a year older decreases the likelihood of diagnosis with one of these conditions by 67 percent. Conditional on

[3] Duncombe and Yinger (2005) detail methods to estimate the extra costs of educating disadvantaged students.
[4] Goodman, Gledhill, and Ford (2003) find a similar negative relationship between relative age and child psychiatric disorders in the United Kingdom.

Case 2:

The 2 arrows

- Migration

- Reappraisal

POLICY!

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

Digital Preservation Implementation Plan

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

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).
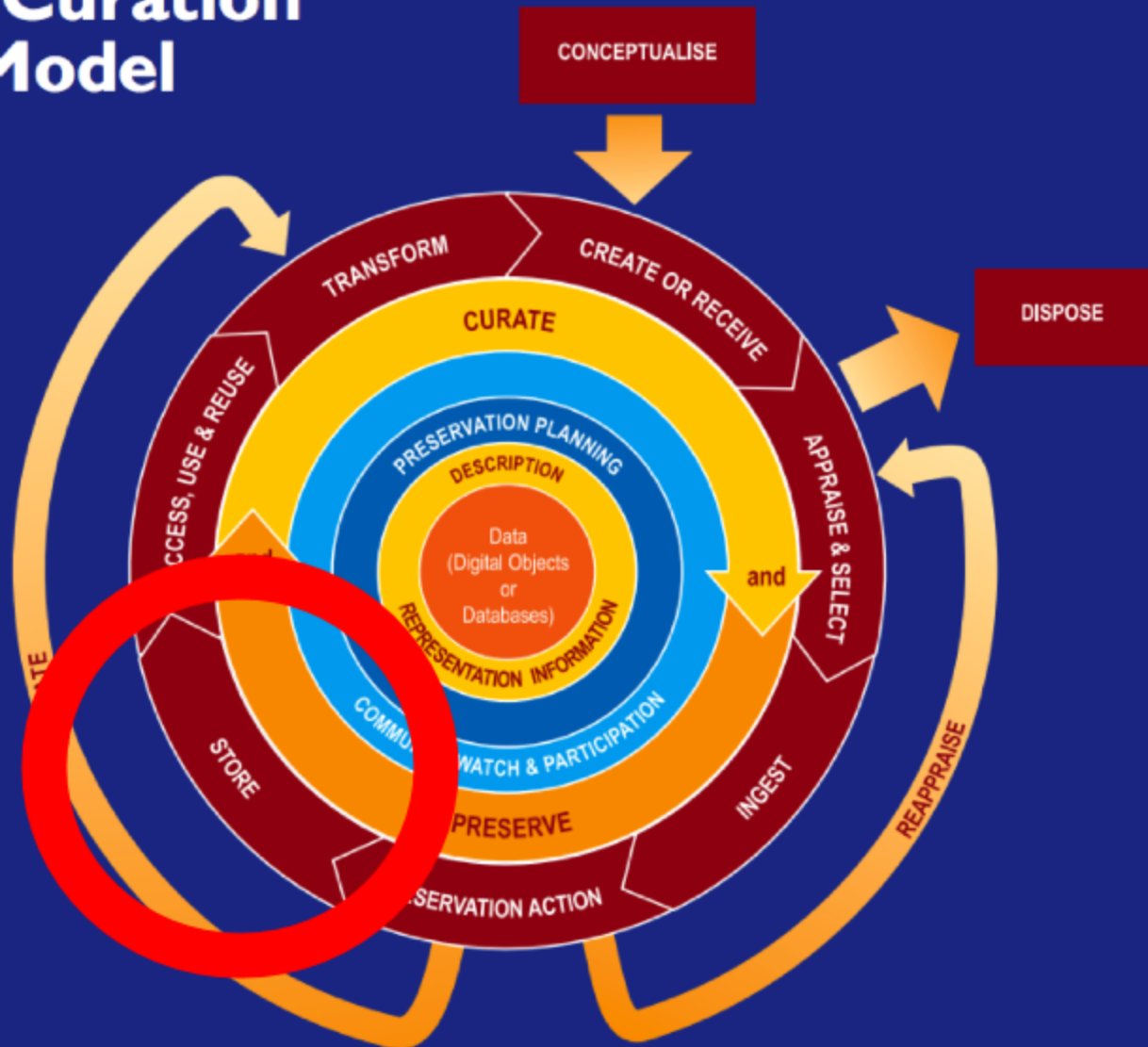
SIP Format

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

The DCC Curation Lifecycle Model

Policy!

backup plan, URIs, file naming plan, fixity

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 folloiwng 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

Tags:
Documentation
Digital Preservation Policy

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 MD5 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

ensure that sufficient description and representation information is stored with data

use a reliable storage medium and geographically distributed backups systems

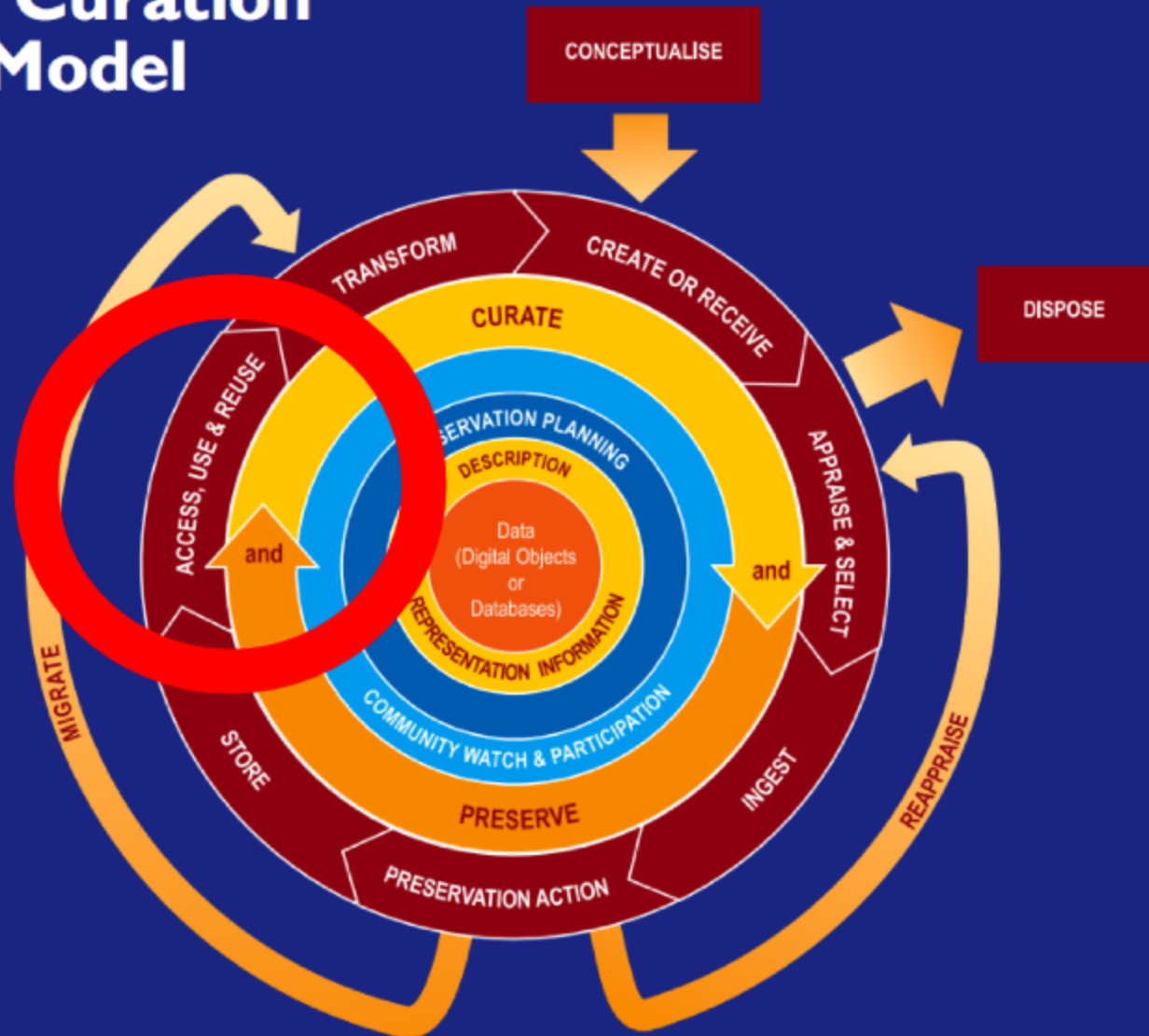**monitor events that might trigger other preservation actions**

regularly check to ensure the integrity of the stored data and their description and representation information

ensure system and physical security

maintain and replace the technical infrastructure as necessary

develop, and administer as necessary, data recovery procedures

The DCC Curation Lifecycle Model

There is **NO** preservation without access.

What's required?

- Appropriate metadata to ensure data can be located.

- Appropriate legal permissions to ensure data can be (re-)used.

- Tools to allow the use of data.

- Access controls.

Policy!

Definition of DIP

Submitted by nruest on Tue, 07/02/2013 - 15:54

Dissemination Information Package (DIP)
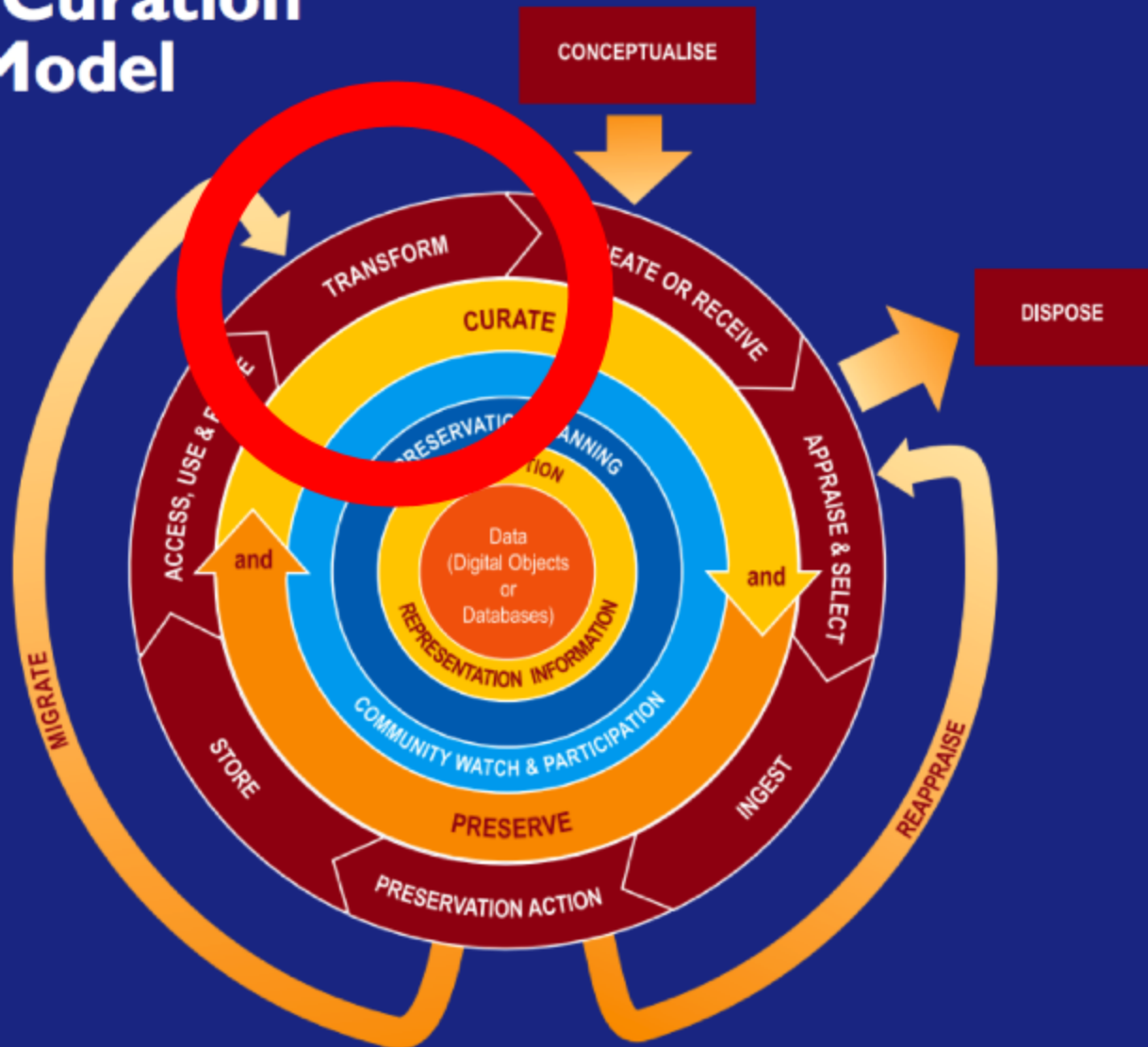
- OAIS describes a DIP as "the Information Package, derived from a part, or all, of one or more AIPs, received by the Consumer in response to a request to the OAIS."
- York University Library's DIPs are always generated from a single AIP.
- User access to archival objects is generally provided through the [YUDL website](http://digital.library.yorku.ca] (http://digital.library.yorku.ca).
- The user is first authenticated, and will, depending on their level of access, see basic object metadata, and a link to the object article as available.
- Context information is provided in the form of links to other items in a given collection.
- The DIP is retrieved using the URI for the corresponding AIP. In turn, the AIP contains metadata tying it back to the SIP.

Tags:

Documentation

Digital Preservation Policy

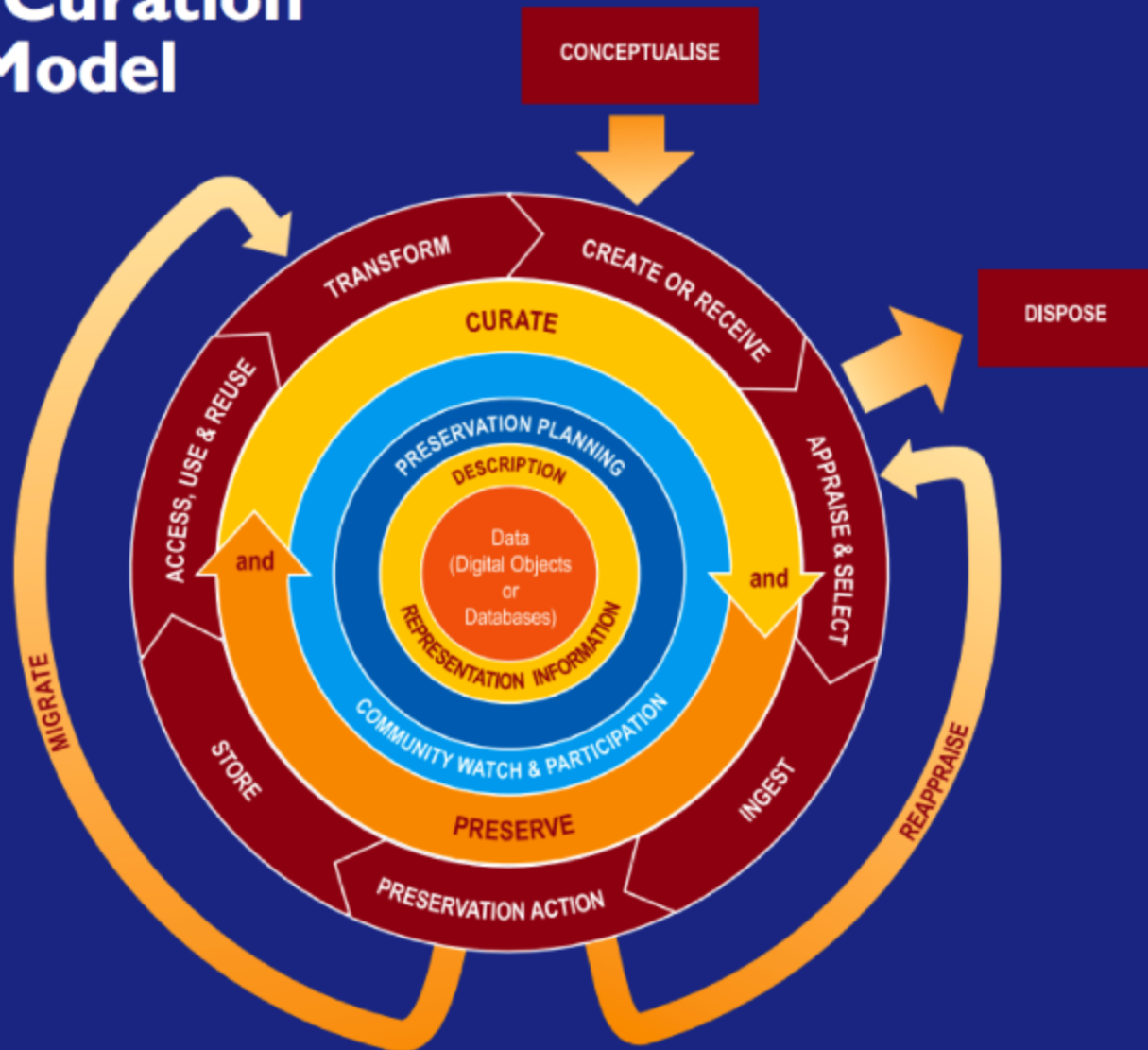The DCC Curation Lifecycle Model

Transform

Transformation

- Can be invoked:
  - At Ingest
  - At time of access
  - As a preservation action

- At ingest: Received formats are not always suitable for preservation
- At access: preservation formats are not always suitable to user needs.
- As preservation: mainly associated with 'Information migration' preservation strategy.

The DCC Curation Lifecycle Model

Community

CurateCamp: digital-curation@googlegroups.com

ALA Digital Preservation: digipres@ala.org

JISC Digital Preservation: DIGITAL-PRESERVATION@JISCMAIL.AC.UK

Digital Preservation Q&A: http://qanda.digipres.org/

The Signal Digital Preservation: http://blogs.loc.gov/digitalpreservation/

spotdocs   Spaces ▾   Browse ▾   Log In

Search

- Digital Curation Bibliography
- Digitization Dreams!
- Magical Ghost
- Meeting Notes
- Member List
- Policies, procedures, and best practices documents
- Product Requirements
- Tool Shed
- Who's Got What Gear?
- Who is Digitizing What?
- Work Plans

OCUL Digital Curation Community
OCUL Digital Curation Community Home

Added by Anika Ervin-Ward, last edited by Nick Ruest on Apr 20, 2014 (view change)

Tools ▾

Terms of Reference

The OCUL Digital Curation Community will bring together OCUL members who have roles in supporting preservation and access to digitized and born digital materials. The scope of this conversation is envisioned as encompassing the full digital curation lifecycle: including digital object creation/digitization, preservation, and the facilitation of access.

The impetus for the OCUL Digital Curation Community comes from the OCUL Digital Curation Summit at York University in May, 2013. There was a clear desire among summit participants for opportunities to share knowledge, experience, work being done at our institutions, and occasions for collaboration. As such, this will be the initial focus of the community. It is also anticipated that the community will serve as a sounding board for issues around digital curation, and provide advice for curation focused activities within OCUL.

As digital curation activities are core to the OCUL mission, this is envisioned as an ongoing community.

Membership

Membership to the OCUL Digital Curation Community is open to all OCUL member institutions. Community membership is at the discretion of each institution and may include more than one person from an institution.

For more information or to become involved contact **Community Moderator**: Nick Ruest, Digital Assets Librarian, York University (ruestn AT yorku.ca).

Note: Any documents that are posted in a members-only area of the OCUL website or SPOT-Docs wiki *shouldn't* be re-posted in in this wiki space.

Members List

Activity Plan

- Work with OCUL membership on opportunities for training in digital curation and preservation. These could be programs led by people in OCUL libraries or by other experts on a particular topic. Areas such as digital forensics and data curation, for example, are taking on importance in digital preservation work; few OCUL institutions have expertise in this area, and arranging training for a single institution may be cost prohibitive.

- Work with OCUL membership on opportunities for shared infrastructure (e.g, platforms and tools such as Islandora, Hydra, Archivematica or methods for cross-institution discovery of digital content). Where support for shared infrastructure is identified, serve as point of contact with OCUL Standing Committees around proposal/implementation details.)

- Work with OCUL membership on consortially negotiated support for digitization of complex formats (e.g., audiovisual recordings, and maps). These formats generally require specialized equipment not found in OCUL institutions. Where support for shared infrastructure is identified, serve as a point of contact with OCUL Standing Committees.

- Hold a reoccurring event focused on digitization and digital curation to provide an opportunity to share knowledge and experience, and to network with colleagues across institutions. This could be a broad, summit-type event, and/or a day focused on a specific topic.

Powered by Atlassian Confluence 5.3.1, Team Collaboration Software · Report a bug · Atlassian News

https://spotdocs.scholarsportal.info/display/ODCC/OCUL+Digital+Curation+Community+Home

Contact

Steve Marks: steve@scholarsportal.info
Nick Ruest: ruestn@yorku.ca