

4-2013

## Doing Digital Preservation at GVSU

Max Eckard

Grand Valley State University, [eckardm@gvsu.edu](mailto:eckardm@gvsu.edu)

Follow this and additional works at: [https://scholarworks.gvsu.edu/library\\_presentations](https://scholarworks.gvsu.edu/library_presentations)



Part of the [Library and Information Science Commons](#)

---

### ScholarWorks Citation

Eckard, Max, "Doing Digital Preservation at GVSU" (2013). *Presentations*. 45.  
[https://scholarworks.gvsu.edu/library\\_presentations/45](https://scholarworks.gvsu.edu/library_presentations/45)

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@GVSU. It has been accepted for inclusion in Presentations by an authorized administrator of ScholarWorks@GVSU. For more information, please contact [scholarworks@gvsu.edu](mailto:scholarworks@gvsu.edu).



# Doing Digital Preservation at GVSU

Save the bits!

# Outline

- Why
- Context
- Reference Models
- Doing Digital Preservation at GVSU
- Speaking Out
- Next Steps
- Conclusion



A close-up photograph of a red brick wall. The bricks are arranged in a standard running bond pattern. The word "WHY?" is written across the wall in a bright blue, hand-painted graffiti style. The letters are thick and slightly irregular, with some drips at the bottom. The word is positioned in the lower-left to center area of the frame.

**WHY**

By Ksayer1

# Why

- Digital preservation is important!
- Library of Congress' Digital Preservation Education and Outreach curriculum, as well as the Heritage Preservation and Association of Southeastern Research libraries are focused on governance and theory
- Replicable for small- to medium-sized institutions



An architectural rendering of a modern building complex. The scene is dominated by a central courtyard area with a winding path and a central water feature. The buildings are multi-story with large glass windows and flat roofs. The landscape is lush with green grass, numerous trees with yellow-green foliage, and a central water feature that flows through the courtyard. The overall atmosphere is bright and modern.

**CONTEXT**

# Context

## Special Collections...

- “...acquires, preserves, and provides access to the University Libraries' growing collection of incunabula and 16<sup>th</sup>-century printing, rare and unique book collections, manuscripts, photographic prints and negatives, maps, broadsides, audio visual materials, and ephemera.”
- “...particular focus on Abraham Lincoln and the Civil War; West Michigan literature, history and culture; and history of the book and printing arts

## Archives...

- “...is the repository for official and quasi-official records created by the University's administration, academic departments, faculty, students, and campus organizations.”
- “...collects records that document and support the University's mission to contribute to the enrichment of society through excellent teaching, active scholarship, and public service.”



**GRAND VALLEY  
STATE UNIVERSITY**

**UNIVERSITY LIBRARIES  
DIGITAL COLLECTIONS**

# Context (cont.)

- Not new to providing access to digital materials
  - Digitizing for quite some time
  - Even accepting born-digital materials
- Pretty new to managing and preserving digital materials
  - We got this. → We're ignoring it. → We can't deal with it. → We can't not deal with it.
  - My position created to help start addressing these issues
- Disclaimer: trusted not Trusted



Analog      digitize      born-digital

~~AT~~ accession      ~~descriptive section~~

export MARC to catalog  
when ready for use

analog  
↓

access & description  
AT (finding aid)

digitize ↓  
process

master → access

ingest (archival)

born-digital  
description finding aid AT

original

if necessary

master

\* if we get compressed  
digitized  
compressed  
or master

master copies  
backed, redundant

# REFERENCE MODELS

access copy

CONTENT (set / instance)  
item level descriptions

Dreamhost Ensemble (A/N)

access

catalog  
@  
item level

use to create

# Reference Models

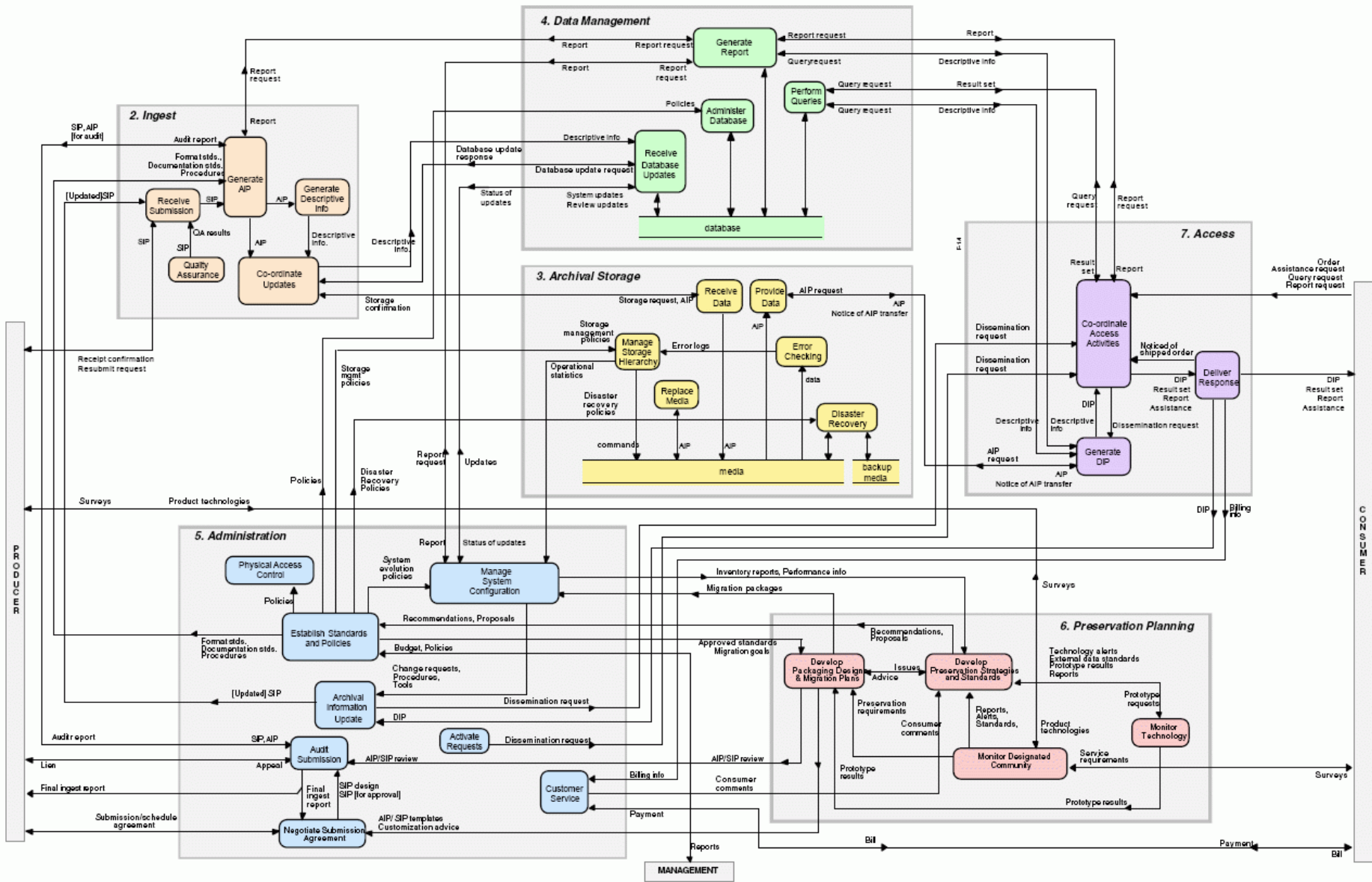
- Open Archival Information System (OAIS)
  - System-centered
- DCC Curation Lifecycle Model
  - Workflow-centered



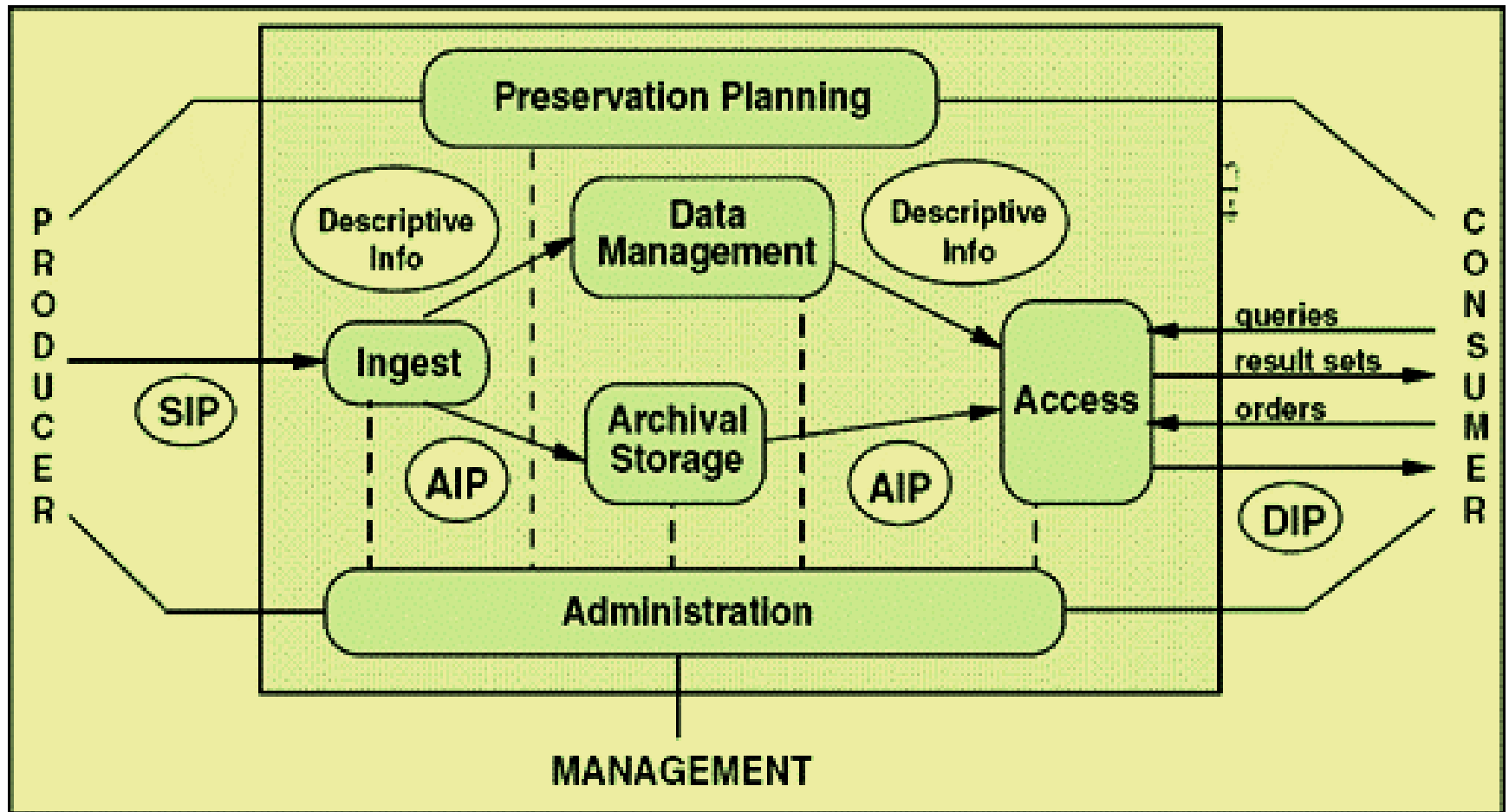
# Open Archival Information System (OAIS)

- Created by Consultative Committee for Space Data Systems (CCSDS)
- Purpose and Scope
  - “...an Archive, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a **Designated Community**.”
  - “The information being maintained has been deemed to need **Long Term Preservation**...”
- Reference Model





# OAIS Functional Entities



# Data and actors model

- “Producers,” “consumers,” and “managers” of data.
- Data + representation = information
  - Surprising how often the representation bit gets lost in translation, but it’s important!
- SIP → AIP → DIP
  - Submission IP: what the producer gives the manager (original)
  - Archival IP: what the manager archives, after any necessary transformation or embellishment (master)
  - Dissemination IP: what consumers see. May or may not be the same as AIP or SIP (access).

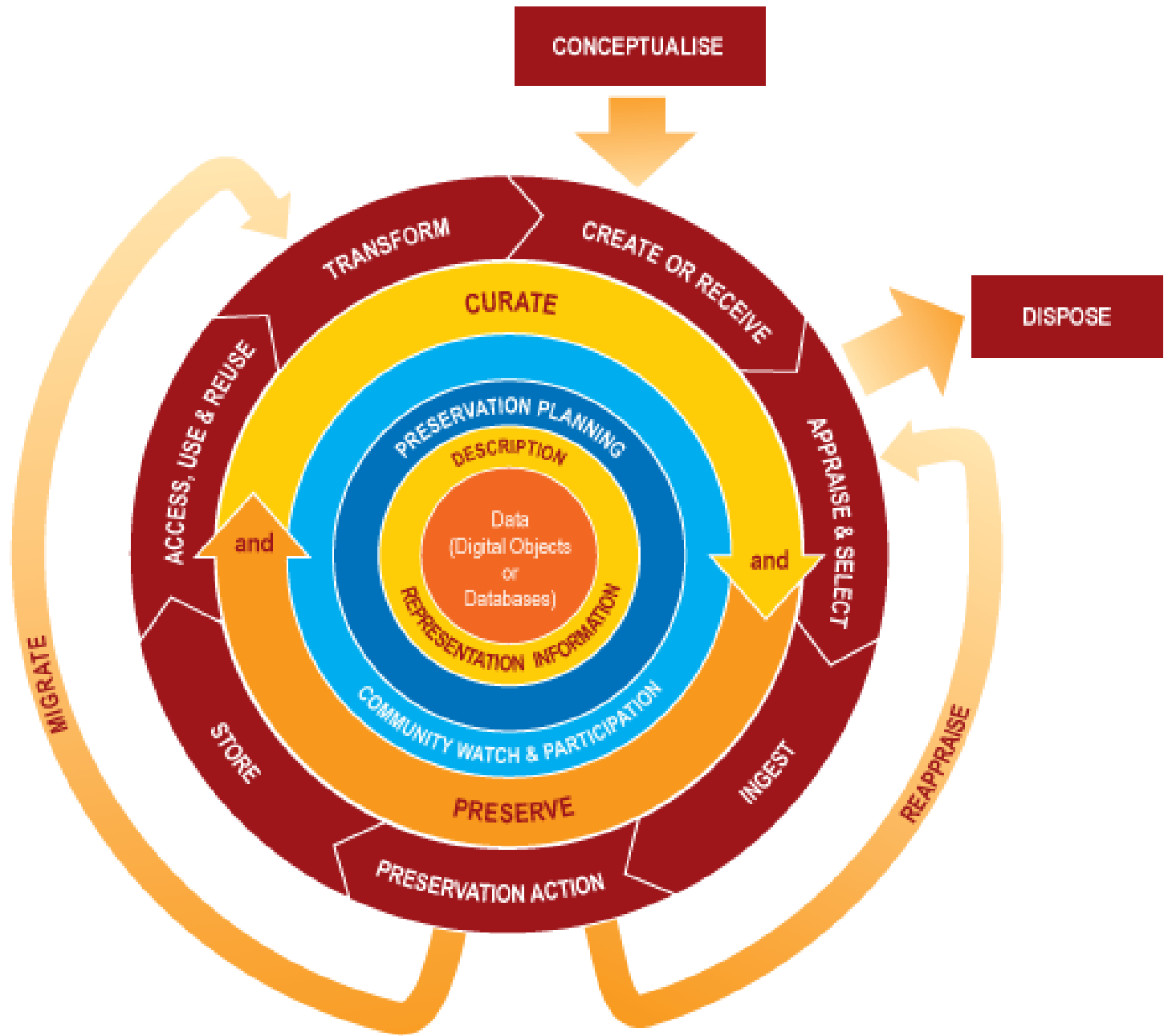


# OAIS requirements, [Dorothea Salo's] version

- Archives must:
  - Get stuff from the people who make stuff.
  - Get their permission to keep and mess with that stuff.
  - Figure out who'll use the stuff.
  - Make sure those users can understand the stuff without needing to bother the people whose stuff it originally was.
  - Write down procedures to preserve the stuff, and follow them.
- That's pretty much it. The rest is commentary. With lots of arrows in it.

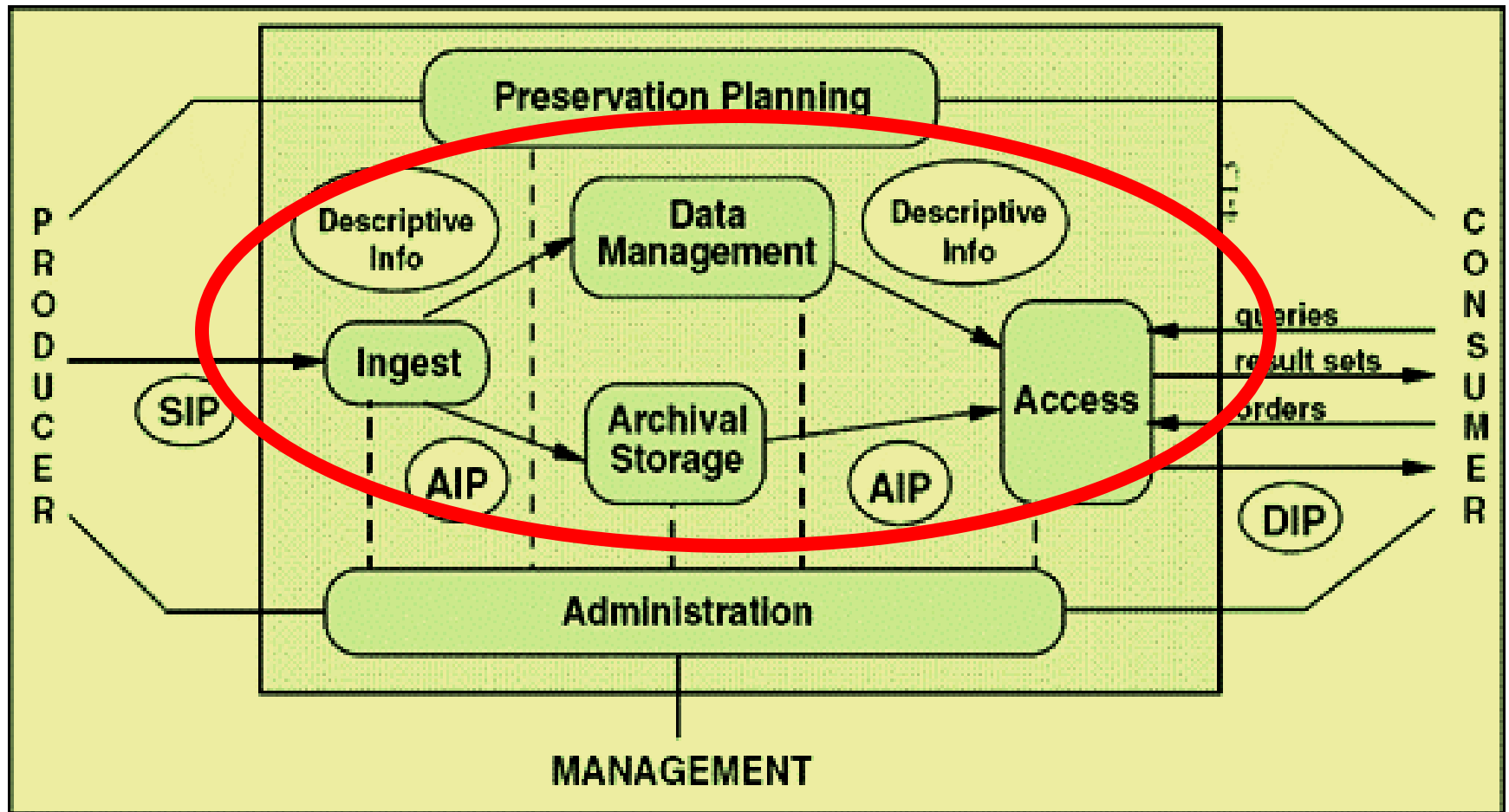
# DCC Curation Lifecycle Model

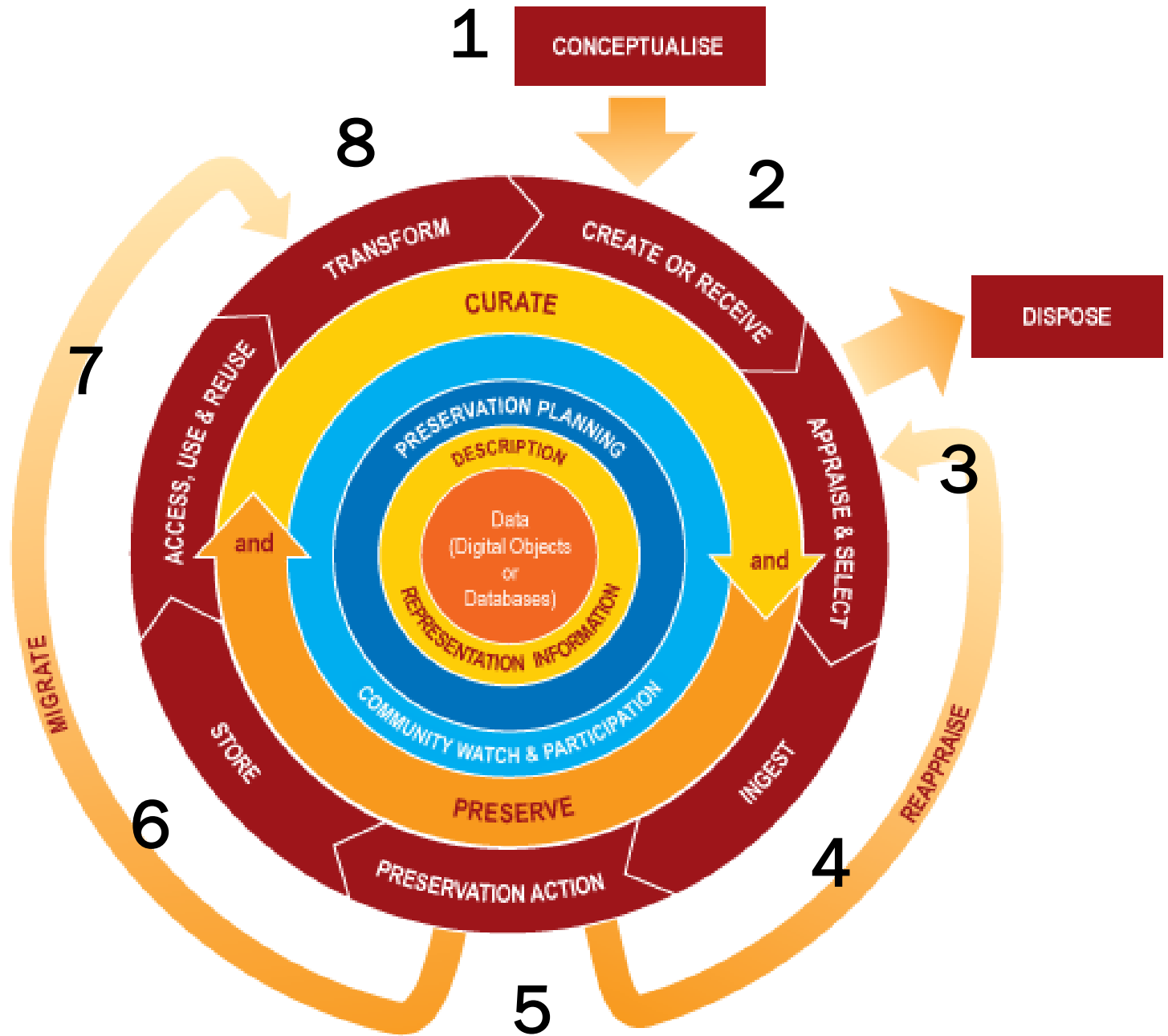
- Digital Curation Centre (DCC)
  - World-leading center of expertise in digital information curation based in the UK
  - Main focus research data management for the UK's higher education research community
- Data = Digital Object(s)





# OAIS Functional Entities







**iPRESERVATIONISTA!**



Digital Access  
Bytes

# DOING DIGITAL PRESERVATION AT GVSU



Save  
the  
Bits





# Conceptualize

## Definition

Conceive and plan the creation of data, including capture method and storage options.

## Tools

- Memorandum of Understanding (MOU)
- Gift Agreement
- Project Plan
  - Digitization
  - Born-Digital

# Project Plan

Contact Information

Project Description

Type/Format/Size of Original Digital Collection

Copyright/Rights/Permissions

Selection Criteria for Proposed Digital Collection

Type/Size of Proposed Digital Collection

Growth Over Time

Description of Proposed Services and Responsibilities/Roles

Notes

## Born-Digital Project Plan GVSU Special Collections

A detailed plan for approval before the born-digital project (a project in which the material has only ever existed digitally) begins:

- Settle terms of Gift Agreement and MOU.
- Determine the goal(s) and scope of the project, determine the details/elements of the metadata, inform all concerned on the workflow, establish responsibilities and timeline.

Today's Date: \_\_\_\_\_

Revised Date: \_\_\_\_\_

### Contact Information

Contact person (proposed by)	
Organization name/dept.	
Address	
Phone	
Website	
Email	

### Project Description

Name of project	
Abstract	
Goals (audience, basic or full preservation, and/or vs. access)	

### + Type/Format/Size of Original Digital Collection

	Format	Size (MB, GB, etc.)
Text		
Images		
Audio		
Video		
Complex Items (websites,		

# Project Plan

Contact Information

Project Description

Type/Format/Size of Original Digital Collection

Copyright/Rights/Permissions

Selection Criteria for Proposed Digital Collection

Type/Size of Proposed Digital Collection

Growth Over Time

Description of Proposed Services and Responsibilities/Roles

Notes

## Born-Digital Project Plan GVSU Special Collections

A detailed plan for approval before the born-digital project (a project in which the material has only ever existed digitally) begins:

- Settle terms of Gift Agreement and MOU.
- Determine the goal(s) and scope of the project, determine the details/elements of the metadata, inform all concerned on the workflow, establish responsibilities and timeline.

Today's Date: \_\_\_\_\_

Revised Date: \_\_\_\_\_

### Contact Information

Contact person (proposed by)	
Organization name/dept.	
Address	
Phone	
Website	
Email	

### Project Description

Name of project	
Abstract	
Goals (audience, basic or full preservation, and/or vs. access)	

### + Type/Format/Size of Original Digital Collection

	Format	Size (MB, GB, etc.)
Text		
Images		
Audio		
Video		
Complex Items (websites,		



# Create or Receive

## Definition

Create data including administrative, descriptive, structural and technical metadata. Preservation metadata may also be added at the time of creation.

*Think digitization projects.*

Receive data, in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata.

*Think born-digital projects.*

## Tools

- Create
  - Digitization
    - Scanners, digital cameras, analog-to-digital converters
    - In-house, in-university, or outsourced
    - Some technical metadata created when digitized that can be harvested later
  - Born-Digital
    - Computers, digital cameras, digital video and audio recorders, cell phones...
    - Metadata: More often hit and miss, since you're not the one in control
- Receive
  - External hard drives, CDs, DVDs, thumb drives, DropBox...
  - Some combination of above





# Appraise and Select

## Definition

Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements.

## Tools

- Collection development policies, mission statement
  - Special Collections
    - Abraham Lincoln and the Civil War
    - West Michigan literature, history and culture
    - History of the book
  - Archives
    - Official and quasi-official records created by the University's administration, academic departments, faculty, students, and campus organizations.
  - Digitizing at-risk analog materials

# Appraise and Select (cont.)

Committed to “saving the bits”

Preserve content

Not necessarily look and feel

Two tiered approach...

## Preservation File Format Guidelines



The University Libraries and Special Collections & University Archives are committed to providing long-term access to all digital content deposited in Digital Collections (and [scollections@gvsu.edu](mailto:scollections@gvsu.edu), Institutional Repository) by applying best practices for data management and digital preservation, while also acknowledging the complexities involved in preserving digital information and the library's existing technical and financial limitations. Staff will preserve the content in the form it is originally deposited and for certain files and at the sole discretion of the staff, staff will attempt to preserve the content, structure, and functionality of the files through migration or other preservation strategies.<sup>1</sup> For all files included in Digital Collections (and [scollections@gvsu.edu](mailto:scollections@gvsu.edu), Institutional Repository) the staff will provide basic services including secure storage, backup, management, file-checks, and periodic refreshment by copying the data to new storage media. To the extent possible, provenance information will be embedded in the files themselves.

- Keep the original bit stream, as well as any "preservation" version in the likelihood that preservation methods are developed in the future.
- Clean and validate data to ensure they can be managed and reused over time.
- Add high-quality preservation metadata and representation information to increase potential for discovery, reuse, and preservation.
- Ensure acceptable data structures or file formats, for example, by using nonproprietary, well-documented data format standards to increase the chance of future recoverability.
- Apply good data management practices.
- Implement secure storage and institutional continuity.

When the University Libraries and Special Collections & University Archives will provide access to digital material, access will be as seamless as possible for all users, regardless of bandwidth, device, browser capabilities or operating system.

OR <http://fileweb.file.edu/node/897>

<sup>1</sup> Digital Collections (and [scollections@gvsu.edu](mailto:scollections@gvsu.edu), Institutional Repository) will make limited efforts to maintain the usability of the file as well as preserving it as submitted (bit-level preservation). The format will be monitored and may be transformed when significant risk to access is imminent but it is likely to be difficult to predictor control the consequences of any transformation or migration on content, structure or functionality. The file may also be transformed to a more "preservation-friendly" format to ensure that the information content is not lost, even if some structure and functionality are sacrificed.

# Appraise and Select (cont.)

## Basic Preservation Support

- Onsite and offsite storage
- Regular backups
- Periodic copying to new storage media (media refreshment)
- Periodic fixity checking
- Periodic verification that the AIP is complete

## Full Preservation Support

- Normalization
- Migration

# Appraise and Select (cont.)

1. Is the content worth saving?
2. Is the original format still created and accessible?
  1. LC Sustainability of Digital Formats site
  2. Florida Recommendations
3. Is the quality of the normalized version as good or better than the original version?
4. Is there content or metadata or function that was not transferred at the time the preservation copy was made? Is the loss significant?
5. Is the space available to retain both copies?
6. Is the cost of retaining both justified?



# Ingest

## Definition

Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance, policies or legal requirements.

## Tools

- Accession: AT
- Make sure you get everything you were supposed to: CloneSpy
- Scan for viruses: Symantec
- Start organizing: FreeCommander
- Get a sense of what you got:
  - Properties
  - Digital Multimedia: Gspot
- Get your file names in order:
  - Renamer
    - Add a unique prefix
    - Remove special characters
- NOTE: All of this happens on local, networked servers backed up by our campus IT department





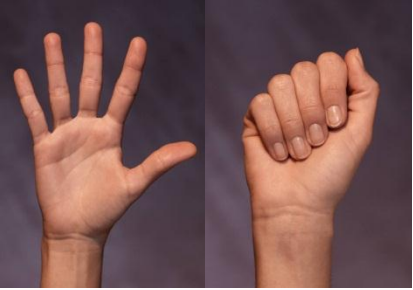
# Preservation Action

## Defintion

Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity.

## Processes

- Add preservation metadata/descriptive info to the original
- Migrate (create a master)
  - Specified directory structure
  - File naming convention
- Add preservation metadata/descriptive metadata to the master



# *Descriptive Info*

## PREMIS

- Provenance, authenticity, preservation activity, technical environment, rights management
- For example:
  - messageDigest Algorithm (1.5.2.1), messageDigest (1.5.2.2)
  - objectCharacteristics (1.5)
  - originalName (1.6)
  - objectIdentifierValue (1.1.2)
  - contentLocationValue (1.7.1.2)

## Tools

- HashMyFiles, MD5summer
- Spreadsheet, database



PRESERVATION METADATA  
MAINTENANCE ACTIVITY

## ***Descriptive Info (cont.)***

Preservation metadata guidelines

Data dictionary

What fields we use and how

- Required, Recommended
- What CVs we use

### Preservation Metadata Guidelines



The University Libraries and Special Collections & University Archives General Metadata Guidelines provide a set of rules for the creation of PREMIS metadata records for preservation.<sup>1</sup> The following guidelines represent the minimum required fields for all SIP and AIP collections in Digital Collections. Additional fields may be added at the discretion of staff.

For digitized collections, preservation metadata should be added to the master copies of files.

For born-digital collections, preservation metadata should be added to the original and master copies of files.

#### PREMIS<sup>2</sup>

##### Message Digest Algorithm, Message Digest

###### Checksum

A form of redundancy check, the checksum can be used to detect errors unseen by the human eye. It does this by adding up the bits and storing the resulting value. The checksum value is a string of alphanumeric characters.

Input guidelines:

1. Generate a checksum using a checksum utility (i.e., [checksum32](#))
2. Specify the specific algorithm used to construct the checksum followed by a colon (i.e., MD5, SHA-1).
3. Record the alphanumeric value as indicated below as it is generated by the utility.

#### Object Characteristics

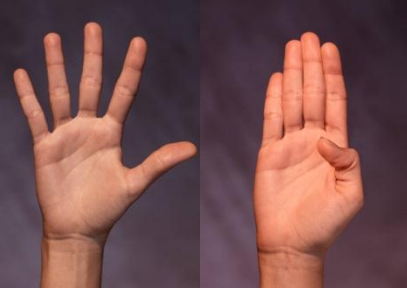
##### Digital Characteristics

The electronic format of the resource being described. For digital assets, format may include the extent of the digital resource, such as file size or playtime.

Input guidelines:

<sup>1</sup> CONTENTdm requires an Object File Name

<sup>2</sup> [http://digital.nyu.edu/cdm4/images/pdf/PreservationMetadataDataDictionary\\_v0v3.pdf](http://digital.nyu.edu/cdm4/images/pdf/PreservationMetadataDataDictionary_v0v3.pdf)



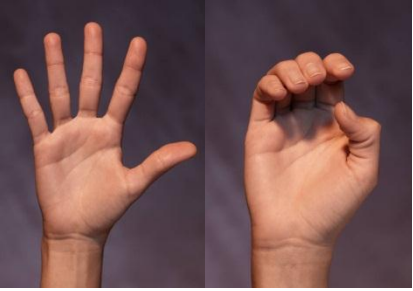
# ***Occasional Action: Migrate/Normalize***

## **Definition**

Migrate data to a different format. This may be done to accord with the storage environment or to ensure the data's immunity from hardware or software obsolescence.

## **Tools**

- Create master/access copies (if necessary):
  - Text (transcripts): Notepad, Notepad++ (free)
  - Word Processing: Open Office (free), Word, Adobe Acrobat Professional (not free)
  - Images: IrfanView (free), PhotoShop (not free)
  - Audio: Audacity, Cdex (free)
  - Video: HandBrake, Miro, VLC Media Player (all free), AVS 4 You, Adobe Premier, Adobe Media Encoder (not free)
  - Caffeine!



# *Preservation Planning*

- Migration and emulation
  - Keep the originals to support migration and emulation preservation strategies.
- Normalization
  - Normalize to file formats with open standards.
  - Formats should be based on best practices and take into account significant properties of file.
- Format policies



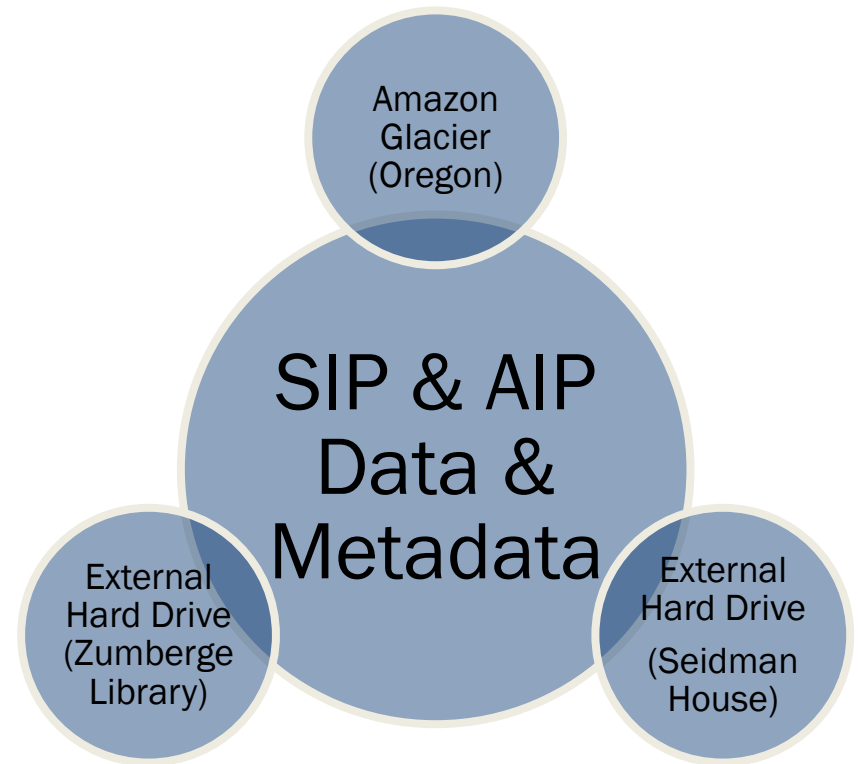


# Store

## Definition

Store the data in a secure manner adhering to relevant standards.

## Methods





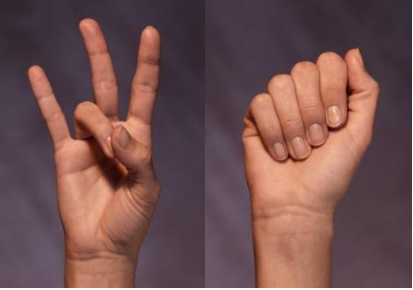
# Access, Use and Reuse

## Definition

Ensure that data is accessible to both designated users and reusers, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable.

## Tools

- Create access copies
  - Store locally for reference
- Resource: AT
- Collection-level
  - Finding Aids
    - EAD → CONTENTdm
    - PDF → In-person researcher
    - MARC → Catalog
- Item-level
  - CONTENTdm
  - YouTube, Flickr
  - Summon
  - Google (60%)
    - SEO
  - Viewshare



# *Descriptive Info*

## Dublin Core

- Title
- Creator
- Description
- Contributor
- Publisher
- Subject
- Identifier (file name)
- Language
- Rights
- Type
- Format
- Date

## Tools

- Spreadsheet, database
- Who does it?
  - In-house
  - Subject experts (i.e., donors)
  - Crowdsourcing!



Dublin Core<sup>®</sup> Metadata Initiative

*Making it easier to find information.*

## Descriptive Info (cont.)

General metadata guidelines

Data dictionary

What fields we use and how

- Required, Recommended
- What CVs we use

### General Metadata Guidelines



The University Libraries and Special Collections & University Archives General Metadata Guidelines provide a set of rules for the creation of Dublin Core metadata records for access and PREMIS metadata records for preservation. The following guidelines represent the minimum required fields for access for all collections in Digital Collections. Unless indicated below, specific implementation guidelines are collection-specific. Additional fields (mapped or unmapped) may be added at the discretion of staff.

#### Dublin Core<sup>1</sup>

##### Mandatory

###### Title

###### Required

**Access given to the resource. Unneeded for SEO.**

###### Best practices:

- Prefer literal and non-numeric description of resource, including material-type information if possible.
- Prefer non-use of explanatory or qualifying symbols (e.g., brackets to indicate catalog-supplied title).
- If the resource has multiple titles (e.g., translated titles, etc.), prefer to use Title-Alternative element.
- Documents with the same title are often considered duplicates by Google.

###### Creator

**Do not be primarily responsible for making the resource.**

###### Best practices:

- Examples of a Creator include a person, and organization, &c.
- "Prefer use of Name (personal or corporate) Authority Source to be used consistently throughout description of a resource and from one resource to another" (Metadata Implementation Guidelines for North Carolina State Digital State Documents).
- Prefer non-use of 'junk value' (e.g., "Unknown,") however, it is appropriate to qualify named entities with "[etc]".

<sup>1</sup> [http://www.sdicons.gatech.edu/support/best\\_practices.pdf](http://www.sdicons.gatech.edu/support/best_practices.pdf)



# Transform

## Definition

Create new data from the original, for example:

- by migration into a different format (*That's us*), or
- by creating a subset, by selection or query, to create newly derived results, perhaps for publication (*That's our users*)

## Products

- Us
  - Change to data or metadata after initial data management process
    - New storage medium, better file formats
    - Revisit policies every year[ish]
    - Keep up with the latest and greatest
- Them
  - Featured in books/magazines/other publications
  - Used in research
  - Used in social media
  - Used in classrooms
  - Used in documentaries
  - To brighten people's day...





# **SPEAKING OUT**

By Superloop

# About

- “...an oral history project dedicated to documenting the history of civil rights and social justice advocacy in Western Michigan.”
- “...this initiative aims to...ensure that these critical histories are preserved for future generations.”
- Areas covered:
  - Race relations & ethnic identity, LGBT issues, immigration, disability/ability, veterans, women and gender (not LGBT)
- 110 open interviews, 148 interviews total





**NEXT STEPS**

By Martin

# Next Steps

- Still feels very manual
  - Automation, batch processing
  - code{4}lib
    - AutoHotKey
  - Ingest:
    - Web-harvesting
    - FITS
    - CINCH
  - Archivematica, Preservica, Rosetta (Ex Libris)
- Rocking the command line
  - ffmpeg
  - GhostScript
- More sophisticated archival storage
  - Sync
  - DuraCloud
- Open source options for DAM
  - DSpace, Fedora, Islandora, &c.



**QUESTIONS?**

By Ninja M.

# Thanks

## Links

### Why

- Digital Preservation Education for NC State Government Employees  
<http://digitalpreservation.ncdcr.gov/index.html>

### Reference Models

- OAIS:  
[http://en.wikipedia.org/wiki/Open\\_Archival\\_Information\\_System](http://en.wikipedia.org/wiki/Open_Archival_Information_System)  
<http://public.ccsds.org/publications/archive/650x0m2.pdf>
- DCC Curation Lifecycle Model:  
<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

### Doing Digital Preservation at GVSU

- FDA Preservation Support Levels  
<http://fclaweb.fcla.edu/content/fda-preservation-support-levels>
- Sustainability of Digital Formats  
<http://www.digitalpreservation.gov/formats/>
- PREMIS  
<http://www.loc.gov/standards/premis/>  
[http://digital.ncdcr.gov/ui/custom/default/collection/default/resources/custompages/about/preservation\\_metadata\\_data\\_dictionary\\_rev5.pdf](http://digital.ncdcr.gov/ui/custom/default/collection/default/resources/custompages/about/preservation_metadata_data_dictionary_rev5.pdf)

- Format policies  
[https://www.archivematica.org/wiki/Format\\_policies](https://www.archivematica.org/wiki/Format_policies)
- Dublin Core  
<http://dublincore.org/>

### Speaking Out

- Speaking Out:  
<http://www.gvsu.edu/speaking/>

### Next Steps

- Issue 20, 2013-04-17  
<http://journal.code4lib.org/issues/issues/issue20>

## Me

Max Eckard

Metadata & Digital Curation  
Librarian

Grand Valley State University

(616) 331-5072

eckardm@gvsu.edu



**GRAND VALLEY  
STATE UNIVERSITY**  
UNIVERSITY LIBRARIES  
DIGITAL COLLECTIONS