FAAQ: Frequently Asked Archivematica Questions

Poll – please go to link or QR code

Anonymous – not collecting email accounts for responses

Questions for the panel will remain open throughout (also anonymous)

https://forms.gle/hLNt7CVjtsbMr11Z7

- Lauren Goodley, Archivist, The Wittliff Collections, Texas State University
  - *"Digital Preservation Overview"*

- Chris Bañuelos

  Digital Archivist, Woodson Research Center, Rice University
  - *"Introduction to Archivematica"*

- Bethany Scott, Head of Preservation and Reformatting, University of Houston Libraries
  - *"FAAQ: University of Houston Libraries Case Study"*

DIGITAL PRESERVATION OVERVIEW

High-Level Concepts

- 3-Legged Stool
- LOCKSS

Deep Dive

- OAIS
- AIPs
- PDI

Actionable Steps

- NDSA Levels of Preservation
- Content
- Fixity/Integrity

TEXAS ★ STATE
UNIVERSITY ®

Digital Preservation Management: Implementing short-term strategies for long-term problems
https://www.dpworkshop.org/dpm-eng/conclusion.html

3-Legged Stool

- **Organizational Infrastructure:** the policies, procedures, practices, people

- **Technological Infrastructure:** equipment, software, hardware, secure environment, and skills

- **Resources Framework:** startup, ongoing, and contingency funding

TEXAS ★ STATE
UNIVERSITY ®

Good Practice vs Best Practice

- It depends

- Unattainable

- Different practice makes sense for different organizations and materials

"When is 'good' better than 'best'? In support of digital preservation good practice," Jenny Mitcham, 18 April 2023, https://www.dpconline.org/blog/when-is-good-better-than-best-in-support-of-digital-preservation-good-practice.

TEXAS ★ STATE
UNIVERSITY ®

Lots Of Copies Keep Stuff Safe

- **Copies in disparate geographical, administrative, and technological locations**
- **Distributed storage**
- **Consortia (DuraCloud@TDL, Amazon, Chronopolis**

Library of Congress Digital Preservation Partners
https://www.digitalpreservation.gov/partners/lockss.html

LOCKSS

- Open-source system of networked data replicas – shared copies of e-journals – that allows the participants to access reliably preserved data.

- CLOCKSS – controlled LOCKSS for archival materials

- Adopted concepts from Stanford/LOC for archival digital preservation – DuraCloud, NDSA Levels
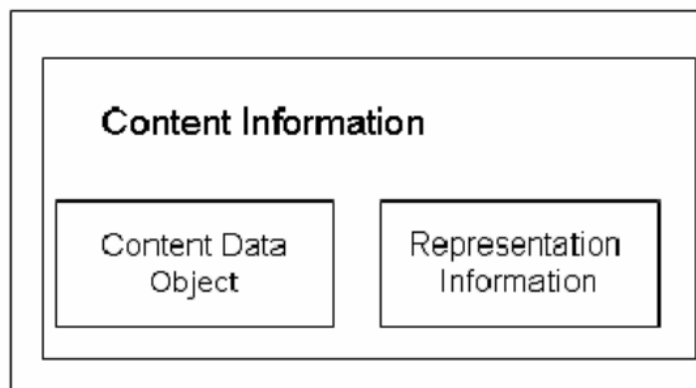
TEXAS ★ STATE
UNIVERSITY ®

OAIS Reference Model
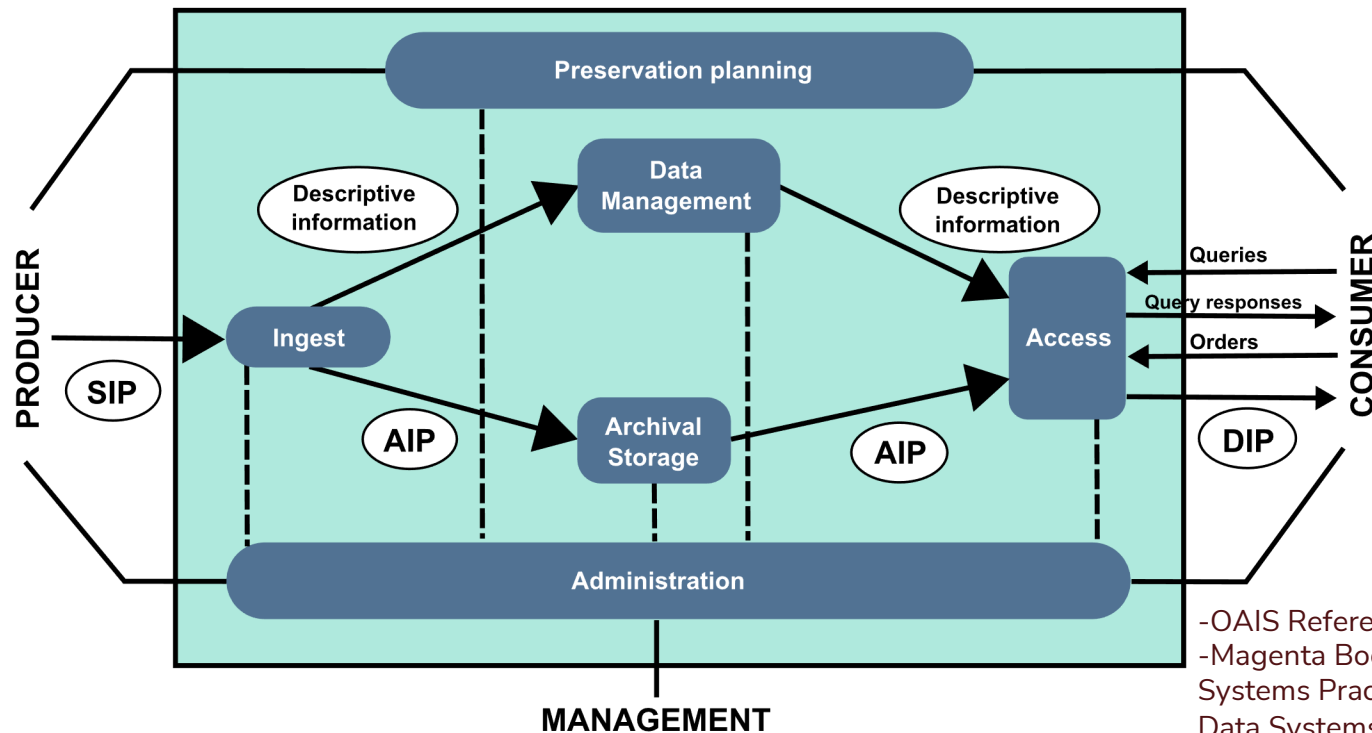
- ISO Standard: ISO 14721
- Open Archival Information System
- Reference Model –tool and policy agnostic
- Organizational Infrastructure – must map and incorporate tools, services, and policies to adhere to this model

-OAIS Reference Model oais.info
-Magenta Book, Recommendataion for Space Data Systems Practices, Consultatative Committee for Space Data Systems https://public.ccsds.org/pubs/650x0m2.pdf

AIP – Archival Information Package Structure

Content Information

Content Data Object	Representation Information

1. Content/Data file
2. Information needed to present that file to designated community

Preservation Descriptive Information

Reference	Provenance	Access Rights
Context	Fixity	
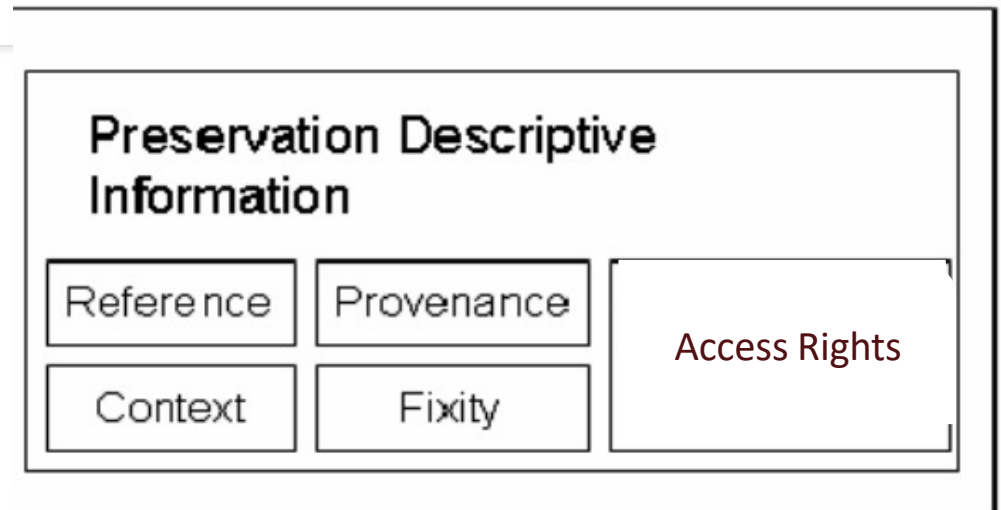
1. Reference— identifier
2. Provenance – history of the item
3. Context – for original publication
4. Fixity – checksum
5. Access Rights

Functional Areas:
Storage
Integrity
Control
Metadata
Content

Levels
1-Know your content
2-Protect your content
3-Monitor your content
4-Sustain your content

NDSA

Levels of Digital Preservation

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

Levels of Digital Preservation v2.0

TEXAS ★ STATE
UNIVERSITY ®

Functional Area

(Level)

Level 1 (Know your content)
Document file formats and other essential content characteristics including how and when these were identified

Content

Archivematica

Provides Microservices
- **Complete this task**
  - FIDO
  - File Extension
  - Siegfried
- **Document task**
- **Save in AIP**

Functional Area	(Level)	Archivematica

Integrity	**Level 1 (Know your content)**
	
	Verify integrity information if it has been provided with the content
	
	Generate integrity information if not provided with the content
	
	Virus check all content; isolate content for quarantine as needed
		Provides Microservices
		• **Complete this task**
		  • Submission Documentation
		  • Assign Checksum
		  • Scan for viruses, isolate or quarantine
		• Document task
		• Save in AIP (Fixity →AIP→PDI)

TEXAS★STATE
UNIVERSITY

Sources

Lauren Goodley
lgoodley@txstate.edu

- Digital Preservation Management: Implementing short-term strategies for long-term problems
https://www.dpworkshop.org/dpm-eng/conclusion.html

- Library of Congress Digital Preservation Partners
https://www.digitalpreservation.gov/partners/lockss.html

- OAIS Reference Model oais.info

- Magenta Book, Recommendataion for Space Data Systems Practices, Consultatative Committee for Space Data Systems https://public.ccsds.org/pubs/650x0m2.pdf

- Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware
Storage. https://www.usenix.org/legacy/event/tapp09/tech/full_papers/factor/factor.pdf

- **Digital Preservation Coalition Wiki,** 4.2.1.4.2 Preservation Description Information,
https://wiki.dpconline.org/index.php?title=4.2.1.4.2_Preservation_Description_Information

- Digital Curation Centre, Glossary, https://www.dcc.ac.uk/about/digital-curation/glossary#P

- Levels of Preservation Revisions Working Group, "Levels of Digital Preservation Matrix V2.0,"
October 2019, https://osf.io/2mkwx/ .

TEXAS ★ STATE
UNIVERSITY ®