Core Issues in Digital Preservation: Storage and Maintenance

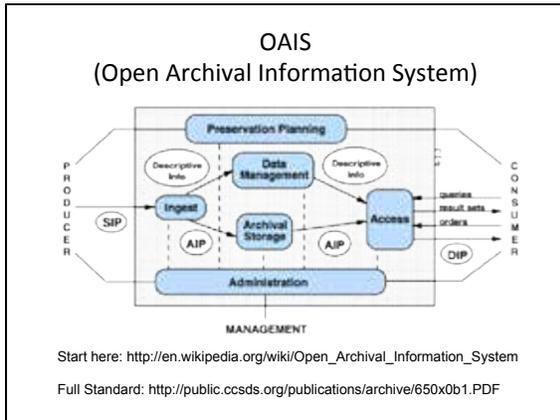Jacob Nadal, Preservation Officer
UCLA Library

Storage and Maintenance

Digital Repositories

- Ingest: Get things in
- Manage: Take care of them
- Disseminate: Get them to users

OAIS
(Open Archival Information System)

Start here: http://en.wikipedia.org/wiki/Open_Archival_Information_System

Full Standard: http://public.ccsds.org/publications/archive/650x0b1.PDF

OAIS v Core IT

- OAIS contains elements that are common in managed IT environments
  - Database administration
  - System Backup
  - Media replacement
- OAIS has concepts that are more specifically archival
  - Preservation Planning and Metadata take a longer view than routine updates and digital asset management
    - "Designated Community" of users determines if archive is usable
  - "Information Packages" as distinct, granular objects
    - In the loosest form of backup, objects may not be handled with the level of independence OAIS expects

OAIS Entities:
Data Management

OAIS Entities: Archival Storage

_____

_____

_____

_____

_____

_____

_____
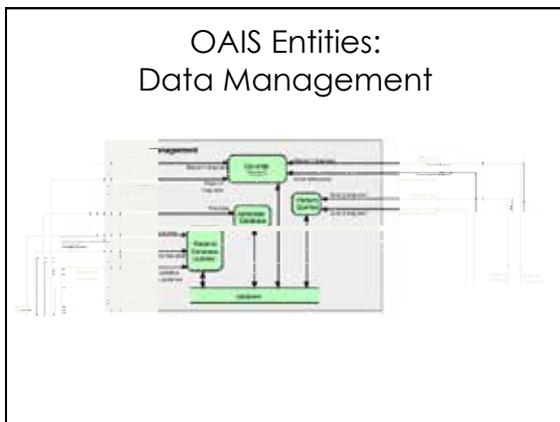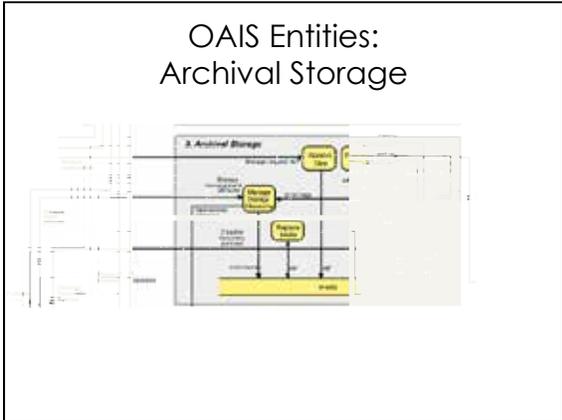
Trusted Digital Repositories

1. OAIS compliance
2. Administrative responsibility
3. Organizational viability
4. Financial sustainability
5. Technological and procedural suitability
6. System security
7. Procedural accountability

_____

_____

_____

_____

_____

_____

_____

Bare bones, or, not *not* digital preservation?

- A 1TB hard drive: $199
- Another 1TB hard drive: $199
- Yet another 1TB hard drive: $199
  – That's $600 for 1TB, very safe, for a year
- Software
  – Text Editor: Pref. w/XML support
  – PDF: Output PDF/A
  – Image: Output TIFF, JPEG 2000; ICC profiles
  – Audio: Output .WAV (Uncompressed PCM)
  – Video: Wait if possible; uncompressed .AVI

_____

_____

_____

_____

_____

_____

_____

Slide 1

	Drive 1 / Workstation	Drive 2	Drive 3	Drive 4
January	Onsite backup	--	--	--
February	Onsite backup	Jan Backup	--	Jan Backup
March	Onsite backup	Offsite	Jan-Feb Backup	Offsite
April	Onsite backup	Offsite	Offsite	Jan-Mar Backup
May	Onsite backup	Jan-Apr Backup	Offsite	Offsite
June	Onsite backup	Offsite	Jan-June Backup	Offsite
July	Onsite backup	Offsite	Offsite	Jan-June Backup
August	Onsite backup	Jan-July Backup	Offsite	Offsite
September	Onsite backup	Offsite	Jan-Aug Backup	Offsite
October	Onsite backup	Offsite	Offsite	Jan-Sept Backup
November	Onsite backup	Jan-Oct Backup	Offsite	Offsite
December	Onsite backup	Offsite	Jan-Nov Backup	Offsite

Slide 2

1. This is not digital preservation, but it is a viable way of getting digitized content through the year.

2. Don't avoid digitization because you don't have a digital repository set up.

3. Don't just keep digitizing or promising long-term preservation without developing (or contracting with) a repository

Slide 3

Not so bare bones

- Fedora Digital repositories
- LOCKSS networks
- DIY Repositories

Not so bare bones

- Fedora digital repositories software:
  - Identifies digital objects
  - Asserts relationships among digital objects
  - Links "behaviors" (i.e., services) to digital objects.
- Open source software: Free to use and develop on your own (http://fedora-commons.org/)
- Also available through a fee-based service called DuraCloud (http://www.duracloud.org)
- Fedora (repository) + D-Space (interface)
  - $4,500-$7,000 / year for .5 TB – 1 TB
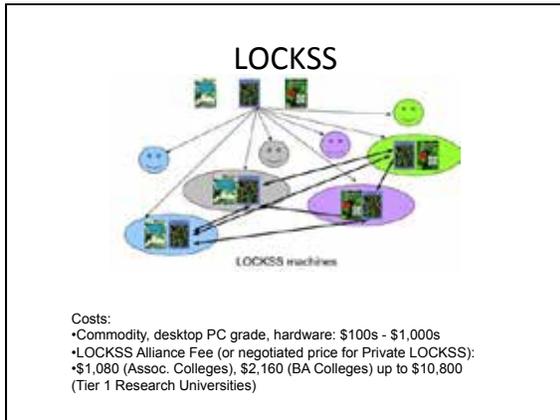 - $1,000/TB per year for extra storage

Fedora

- Kahn and Wilensky Framework
  - www.cnri.reston.va.us/k-w.html
- Supports RDF – "semantic triples"
  - [1]Object [2] described by [3] metadata
  - [1] Page image [2] is part of [3] eBook
- Triples relate well-defined, persistently identified bitsreams or "digital objects"

Distributed Storage

"...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident."

--Thomas Jefferson to Ebenezer Hazard, Philadelphia, February 18, 1791.

LOCKSS

LOCKSS machines

Costs:
•Commodity, desktop PC grade, hardware: $100s - $1,000s
•LOCKSS Alliance Fee (or negotiated price for Private LOCKSS):
•$1,080 (Assoc. Colleges), $2,160 (BA Colleges) up to $10,800
(Tier 1 Research Universities)

Private LOCKSS

- Institutions form a mutual-aid system to maintain each other's content
- MetaArchive (www.metaarchive.org/)
- Alabama Digital Preservation Network (www.adpn.org/)
- Other private networks: (www.lockss.org/lockss/Private_LOCKSS_Networks)

Third-party services

- OCLC Digital Archive
  – ContentDM
  – http://www.oclc.org/digitalarchive/
- Cloud Services
  – Currently $500 - $1,000 / TB per year
  – Some level of on-your-own software development
  – Example: http://aws.amazon.com/s3/
  – Example: http://www.sdsc.edu/services/ StorageBackup.html
- Commercial Data Centers

Pros & Cons of Outsourcing

- Pay for what you need, when you need it ("scalable storage")
- Pay for overhead and common denominator services
- Reduces the need for some kinds in-house expertise, and people are expensive
- You need to make a connection between the repository and your access system

Cornell/ICPSR
Digital Preservation Management Framework

http://www.jacobnadal.com/247

STORAGE AND MAINTENANCE Q&A