

Digital Preservation

8/7/2012

Karen Estlund
Head, Digital Library Services
kestlund@uoregon.edu

Threats to Preservation

- Legal Issues and Permissions
 - File Integrity / Bit rot
 - System Failure
 - External Attacks (Hacking, Viruses)
 - Format Obsolescence
 - Media Obsolescence
 - Containers (Inseparable Content)
 - Lack of Context / Information
 - User Error
 - Natural Disasters
-



Key Services of Digital Preservation

- Levels of service, agreements, and policies
 - Commitment to long-term viability
 - Participation in a community of practice & standards
 - Persistent identifiers
 - Data integrity practices
 - Risk management (e.g. redundancy)
 - Discoverable metadata
 - Methods of access
-



Levels of service / policies

- Prioritize by:
 - Risk to originals
 - More important to preserve materials that are more vulnerable (preserving physical copies; born digital items)
- Policies:
 - Review
 - Content: Are there materials that need to be preserved only for a limited amount of time?
 - Container: File formats for obsolescence
 - Technology: risk management
 - Succession plan



Organizational Commitment



LIBRARIES
University of Oregon

Preservation Format Standards

Media	File Format	Specifications
Video	.mp4	High profile
Audio (nature, concert, etc.)	.wav	
Image	.tif	300-600 ppi (16 bit grayscale or 24 bit color)
Image Text (printed)	.tif or .pdf	200-300 ppi
Image Text (handwriting)	.tif	600 ppi
Maps	.tif	600 ppi (shorter than 36"); 300-400 ppi (longer than 36")
Slides	.tif	3000+ ppi



Persistent Identifiers

Uniform Resource Identifier (URI)

- HDL
 - <http://hdl.handle.net/1794/12242>
 - ARK
 - <http://nwda-db.orbiscascade.org/findaid/ark:/80444/xv14723>
 - DOI
 - [doi:10.5061/dryad.60699](http://dx.doi.org/doi:10.5061/dryad.60699)
-



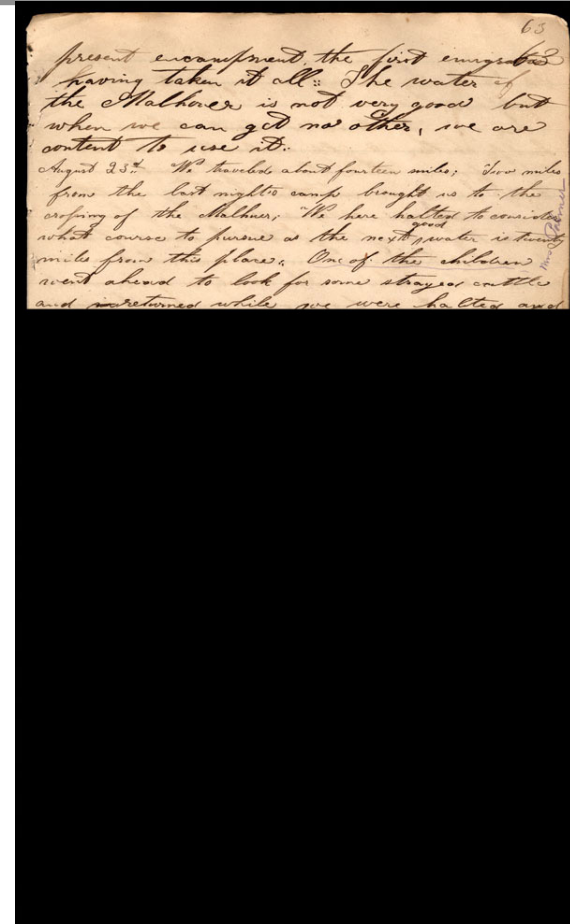
Data Integrity & Risk Management

- File integrity checks
- Refresh
- Back-up
- Off-site storage
 - Preferably in another geographic region
- Mirrored site



Copying as a means of preservation

- Average Server lasts 5 years
- Average Optical Media lasts 6-8 years (CD, DVD)
- Digital Back-up Tapes (LTO) lasts 30 years in ideal conditions



Preservation copies

- Create Digital Object
- Transfer to Editing Tool
- **Export Preservation Format**
- Additional editing/
processing, if necessary
- Export Access Format(s)



DO NOT UPSAMPLE!



LIBRARIES
University of Oregon

Metadata & Access

- Context is important
 - Metadata about the item
 - Metadata embedded in the object
 - File name
- Digital provenance
- Create plan for access
 - Dark, Mediated, Open?



File Naming conventions

- Unique identifier for each file
- Start with collection name, location, identifying information
- Use application-specific 3-letter file extension and lowercase: tif, jpg
- When using sequential numbering, make sure to use leading zeros to allow for multi-digit versions. For example, a sequence of 1-10 should be numbered 01-10; a sequence of 1-100 should be numbered 001-010-100.
- No special characters: & , * % # ; * () ! @ \$ ^ ~ ' { } [] ? < > -
- Only one period and before the file extension (e.g. coll_r.doc
NOT coll0021.b35.f02.001.tif OR coll001_b0956_f0035_d001_p01..tif)
- Try to keep file and folder names under 32 characters
- Keep track of versions (version control)
 - Use a sequential numbered system: v01, v02

E.g. coll001_b0956_f0035_d001_p01.tif



LIBRARIES
University of Oregon