Digital Asset Protection

ALLIANCE FOR RESPONSE – CONCURRENT SESSION
APRIL 27, 2016

Elizabeth La Beaud, MLIS, DAS

DIGITAL LAB MANAGER, THE UNIVERSITY OF SOUTHERN MISSISSIPPI

Roadmap

- Definitions
- Digital Preservation Fundamentals
- Common Preparedness Actions
- Disaster Specific Concerns

Deliverables

- Define and identify your digital assets
- Identify multiple risks to digital assets
- Walk away with *at least* 1 actionable item to improve your digital asset protection strategy.

What is Digital Asset Management?

DIGITAL ASSET MANAGEMENT

- Systems designed to organize and display digital content produced in a variety of media types. The content is usually locally owned and controlled, rather than licensed from a third party.

ELECTRONIC RESOURCE MANAGEMENT

- Systems developed to assist librarians in the control of licensed third-party resources published electronically (databases, e-books, e-journals, etc.)

Examples of Digital Assets

- Digitized photographs
- Word Documents
- Video
- Email
- Websites

Risks to Digital Assets

A Sampling…

- Software obsolescence
- Loss of power
- Bit rot
- Hardware failure
- Cyber attacks
- Format obsolescence
- Media failure or obsolescence
- Man-made and natural disasters
- Dust
- User error

Questions to consider…

- Is the equipment needed to access your assets covered in your insurance plan?
- If your office lost power, could you access your needed files?
- If your desktop disappeared, could you access your needed files?
- If the fire alarm went off, could you leave without stopping to retrieve your server?

Levels of Digital Preservation

NDSA

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

Storage & Geographic Location

Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems

Storage

- Solid State Drives
- Hard Drives
- Optical
- Magnetic

Storage Trends

Why CDs are NOT long term storage…

Storage Trends

Tape → Optical → Solid State → Cloud

Cloud

- Vendor hosted
- Where is your data?
- What happens if your vendor goes out of business…what happens to your data?
- If you don't know, call your vendor and check your contract.

Where the cloud really is…

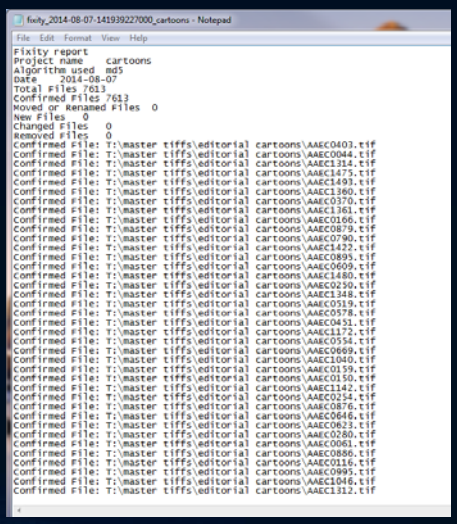
Amazon Web Services

File Fixity & Data Integrity

File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies

Cryptographic Hashes

- Checksums
  - SHA256
  - MD5
- MD5 Summer
- AVPreserve's Fixity
- File Verifier ++
  - (these are windows examples)



Information Security

Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs

Information Security

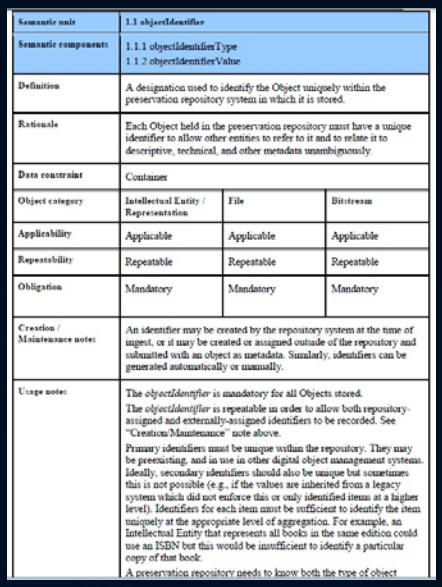
- Make sure no one person has write privileges to all copies.

Metadata

Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata

Metadata

- Descriptive Metadata
  - Dublin Core
  - Mods
- Technical Metadata
  - MIX
- Structural Metadata
  - METS
- Preservation Metadata
  - PREMIS



Semantic unit	1.1 objectIdentifier		
Semantic components	1.1.1 objectIdentifierType 1.1.2 objectIdentifierValue		
Definition	A designation used to identify the Object uniquely within the preservation repository system in which it is stored.		
Rationale	Each Object held in the preservation repository must have a unique identifier to allow other entities to refer to it and to relate it to descriptive, technical, and other metadata unambiguously.		
Data constraint	Container		
Object category	Intellectual Entity / Representation	File	Bitstream
Applicability	Applicable	Applicable	Applicable
Repeatability	Repeatable	Repeatable	Repeatable
Obligation	Mandatory	Mandatory	Mandatory
Creation / Maintenance notes	An identifier may be created by the repository system at the time of ingest, or it may be created or assigned outside of the repository and submitted with an object as metadata. Similarly, identifiers can be generated automatically or manually.		
Usage notes	The *objectIdentifier* is mandatory for all Objects stored. The *objectIdentifier* is repeatable in order to allow both repository-assigned and externally-assigned identifiers to be recorded. See "Creation/Maintenance" note above. Primary identifiers must be unique within the repository. They may be preexisting, and in use in other digital object management systems. Ideally, secondary identifiers should also be unique but sometimes this is not possible (e.g., if the values are inherited from a legacy system which did not enforce this or only identified items at a higher level). Identifiers for each item must be sufficient to identify the item uniquely at the appropriate level of aggregation. For example, an Intellectual Entity that represents all books in the same edition could use an ISBN but this would be insufficient to identify a particular copy of that book. A preservation repository needs to know both the type of object		

File Format Sustainability

File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

File Format Sustainability

- Disclosure
- Adoption
- Transparency
- Self-documentation
- External dependencies
- Impact of patents
- Technical protection mechanisms


THE BENEFITS AND RISKS OF THE PDF/A-3 FILE FORMAT FOR ARCHIVAL INSTITUTIONS

AN NDSA REPORT

February 2014

File Naming Conventions

- No spaces
- Avoid special characters
- Consistent date format
  - YYYY-MM-DD


Why should I use a File Naming Convention ?

A file naming convention (FNC) can help you stay organized by making it easy to identify the file(s) that contain the information that you are looking for just from its title and by grouping files that contain similar information close together. A good FNC can also help others better understand and navigate through your work.

Consider the following examples:

Files without employing an naming convention:

- Test_data_2013
- Project_Data
- Design for project.doc
- Lab_work_Eric
- Second_test
- Meeting Notes Oct 23

Files with a naming convention:

- 20130503_DOEProject_DesignDocument_Smith_v2-01.docx
- 20130709_DOEProject_MasterData_Jones_v1-00.xlsx
- 20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx
- 20130825_DOEProject_Ex1Test1_Documentation_Gonzalez_v3-03.xlsx
- 20131002_DOEProject_Ex1Test2_Data_Gonzalez_v1-01.xlsx
- 20141023_DOEProject_ProjectMeetingNotes_Kramer_v1-00.docx

The files with a naming convention provide a preview of the content, are organized in a logical way (by date yyyy-mm-dd) identify the responsible party and convey the work history, unlike the files without a naming convention.

What would you do in the event of…

- A power outage?
- A flood or leak?
- A fire?
- A tornado?
- A hurricane?
- An earthquake?
- A volcano?

Action List Ideas

- ❑ Non co-location of inventory
- ❑ Call insurance company to see if vital equipment is covered
- ❑ Call vendor to see where your data is stored and what happens to it if vendor goes out of business
- ❑ Pick computers off the floor
- ❑ Purchase UPS
- ❑ Replicate data offsite
- ❑ Migrate files to sustainable formats
- ❑ Adopt sustainable file naming conventions
- ❑ Create list of operational needs
- ❑ Add digital assets into Disaster Recovery Plan

Further Reading…

- Levels of Digital Preservation
  http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf
- File naming conventions
  http://guides.lib.purdue.edu/c.php?g=353013&p=2378293
- Benefits and risks of PDF/A-3
  http://lcweb2.loc.gov/master/gdc/lcpubs/2013655115.pdf
- Hathitrust Disaster Recovery Plan
  http://lcweb2.loc.gov/master/gdc/lcpubs/2013655115.pdf

Questions?