MANAGING DIGITAL CONTENT OVER TIME

Part 3. Store

Digital Preservation Outreach and Education (DPOE)
Identify - what digital content do you have?
Select - what portion of that content will be preserved?
Store - what issues are there for long term storage?
Protect - what steps are needed to protect your digital content?
Manage - what provisions are needed for long-term management?
Provide - what considerations are there for long-term access?
A. Storage needs
B. Well-managed collections
C. Object-level metadata
D. Storage considerations
E. Preservation repository selection
F. Outcomes
A. STORAGE NEEDS

- Think of your objects as packages of data
  - Archival storage manages multi-part content as single objects
  - Digital content (files + metadata = object)
Think of your objects as packages of data
  - Requires some identification and description
    - Metadata
  - Needs at least two copies at least two places
B. WELL-MANAGED COLLECTIONS

- Sample characteristics:
  - Basic information about each deposit
    - *Administrative metadata*
  - Some metadata for objects (you define)
    - *Descriptive metadata*
  - Common (or normalized) file formats
    - *But keep those originals*
  - Controlled and known storage of content
    - *Where is this stuff going?*
  - Multiple copies in at least 2 locations
    - *Where else is this stuff going?*
B. WELL-MANAGED COLLECTIONS

- Basic information about each deposit
  - "Easy" preservation metadata scheme: www.ncecho.org/dig/pmdo.shtml
- Some metadata for objects
  - Descriptive metadata
- Common (or normalized) file formats
  - TIFF, JPG, PDF
  - Free image editor: www.irfanview.com
  - But keep those originals
- Controlled and known storage of content
  - Where is this stuff going?
- Multiple copies in at least 2 locations
  - Where else is this stuff going?
C. OBJECT-LEVEL METADATA

- Metadata enables long-term preservation
  - uniquely identifies digital objects
  - makes digital objects understandable
  - allows objects to be tracked over time
C. OBJECT-LEVEL METADATA

Preservation Metadata

Content (what), Fixity (unchanged), Provenance (life story), Reference (this thing), Context (relationships)

Administrative (manage)

Structural (understand, use)

Descriptive (find, use)

Diagram courtesy DPM Workshops
C. OBJECT-LEVEL METADATA

1. Content: preserve the substance
   - Save the original file, even if you migrate
2. Fixity: demonstrate content is unchanged
   - Checksums: http://www.nirsoft.net/utils/hash_my_files.html
3. Format validation: ensure that it is what it purports to be
   - Jhove: hul.harvard.edu/jhove/windows_xp.html

C. OBJECT-LEVEL METADATA

4. Authenticity: trace to its origin, deposit, and/or your management actions
   - Metadata: Date created, date changed, responsible party

5. Context: preserve linkages with other objects
   - Metadata: Unique identifiers

D. STORAGE CONSIDERATIONS

- Redundancy: How many copies?
  - Minimum: two (2) copies in two location
  - Optimum: six (6) copies

- Other storage questions
  - Are your files too large to store 6 copies? [Videos]
  - Online, near-line, offline?
  - Any legal restrictions? [off-site locations]
  - On what types of media to store your content?
    - Mirrored, networked servers
    - Networked server
    - Your computer and another
    - External hard drives
D. STORAGE CONSIDERATIONS

Factors to consider
- Cost (available resources for preservation)
- Quantity (size and number of files)
- Expertise (skills required to manage)
- Partners (achieving geographic distribution)
- Services (outsourcing)
D. STORAGE CONSIDERATIONS

- Multiple, geographically distributed copies
- Storage partners
- Hosted services
Questions to ask when deciding to use (build, join, buy) a repository

- Is the repository best suited to general or specialized content?
- Do you want an open source or proprietary system?
- How easy is it to manage? [Installation, update, batch upload, etc.]
- Dark or open archive?

Each option has pros and cons

No system is fully compliant to standards

Select best option for your content – for now
Digital preservation requires an organization to:

- Develop a storage management policy
  - E.g., number of copies, locations, fixity means
- Specify storage service or partner agreements
  - Will you give back a fully functioning file in 50 years, or only promise to manage the bits?
- Monitor copies of content for errors/change
- Plan for media replacement
- Consider file obsolescence and how you’ll manage it
  - Migrate and toss or migrate and keep?
F. ORGANIZATIONAL REQUIREMENTS

1. Develop a storage management policy
   - E.g., number of copies, locations, fixity means
2. Specify storage service or partner agreements
   - Will you give back a fully functioning file in 50 years, or only promise to manage the bits?
     - Do you require your storage service to do the same?
3. Monitor copies of content for errors/change
   - FITS: http://code.google.com/p/fits/
   - LOCKSS, MetaArchive
4. Plan for media replacement ($$
5. Consider file obsolescence and how you’ll manage it
   - Migrate and toss or migrate and keep?
TOOLS

- DIGITAL PRESERVATION POLICY TOOL – erpa guidance
  http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf

- HASHMYFILES – Checksum creator
  http://www.nirsoft.net/utils/hash_my_files.htm

- JHOVE – Format validation and identification
  http://hul.harvard.edu/jhove/windows_xp.html

- IRFANVIEW - Free image editor (normalization)
  http://www.irfanview.com

- NC-PMDO - “Easy” Preservation Metadata Element Set
  http://www.ncecho.org/dig/pmdo.shtml