Caring for Digital Materials: Goals

1. Participants will have a better understanding of the inherent fragility of digital objects.

2. Participants will acquire information to help them select preservation formats, metadata, and backup systems for digital objects.

3. Participants will be able to identify one or more actions that can be taken to improve their institution's digital preservation efforts.

Caring for Digital Materials Sessions

We are here:	Overview of digital preservation	Lauren Goodley	Tues., April 2, 2013 2:00 – 3:30 EDT
	Convert it to preserve it: Digitization and file conversion	Jacob Nadal	Thurs., April 4, 2013 2:00 – 3:30 EDT
	Describe it so you can find it: Metadata, finding aids, and asset management	Danielle Plumer	Tues., April 9, 2013 2:00 – 3:30 EDT
	Practice safe archiving: Backups, copies, and what can go wrong	Jefferson Bailey	Weds., April 10, 2013 2:00 – 3:30 EDT
	Partner to preserve: Digital preservation networks and collaboration	Liz Bishoff and Tom Clareson	Mon., April 15, 2013 2:00 – 3:30 EDT

Overview of Digital Preservation

Where to start?

   The basics: concepts, steps, decision points, and considerations

I won't talk about:

   Tools, procedures, answers

Library of Congress Digital Preservation Outreach and Education (DPOE)

DPOE Mission

"The mission of the Digital Preservation Outreach and Education (DPOE) program of the Library of Congress is to **foster** national outreach and education to **encourage individuals and organizations to actively preserve their digital content**, building on a collaborative network of instructors, contributors, and institutional partners."

Today's Topics- 6 Modules

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?

Today's Objectives

★ Have an understanding of digital content management stages.

★ Be able to suggest and implement concrete steps for each stage.

★ Be prepared to get the most out of following workshops.

Module 1

★ Identify - *what* digital content do you have?

Why do we identify content?

- Preservation requires an explicit commitment of resources
- Effective planning is based on knowing the extent of what will be preserved
  - (storage, technology and tools, expertise and knowledge, $)
- Identifying content is a first step to planning for current and future preservation needs
- Not all digital content in and around an organization will be preserved

An explicit inventory is the best way to identify content

Inventory Considerations

★ Inventory content is more important than style and format.

Excel Sheet

Inventory Considerations

★ Inventory content more important than style and format.

★ Inventory results should be:
  ★ Documented: an inventory needs to be captured
  ★ Usable: simple format to sort, list, etc.
  ★ Available: accessible to team, managers, others
  ★ Scalable: content will be added, iterative
  ★ Current: update periodically (iterative) (dated)

Inventory Scope

- What content are we already preserving?
- What other digital content do we have?
- What content do/will our producers create?
- What content are we required to keep?
- What content do we need to review?

Inventory Scope

* Documentation--
  * Retention Schedule
  * Finding aids, catalog
* Projects—
  * Genealogy
  * Digital
* Places—
  * Optical discs—dvds, cds; magnetic discs—old floppies
  * Server
  * Hard drive, desktop, removable hard drive, flash drive

Content Categories

Inventories should include all relevant, e.g.:

* Institutional records
* Special collections materials
* Scholarly content – licensed and open
* Research data
* Web content

Level of Detail

* Inventories can be general to detailed
* Determine appropriate level of detail for you
* Factors in determining level of detail:
  * Extent of content to be inventoried (level)
  * Nature and location of content to be inventoried
    * Off-site? Already described? Media?
  * Resources available to complete inventory
  * Timeframe deadlines for completing inventory

Content Types

* Yearbooks
* Photographs
* Licensed/purchased
* Records schedule materials
* Special collections
* Web Pages
* Genealogy

Format Types

* Born digital
* Audio/video
* Photographs/images
* Text
* Structured data
* Fragile analog
* Maps

What information should you collect?

* Format
* Content
* Date or date range
* Location—physical or file location
* Size (folders or files)

Tip: Dates

Inventories should note:

- Date of inventory – and updates to it
- Date of files – when possible
- Dates covered in content – even approximate
- Date created/received – if relevant, possible

Tip: Location

Locations of content are important – consider:

- Specify online/offline location
- General location – e.g., with us, with creator
- May need to change locations as content moves

Be clear enough without going to extremes…

Module 1: Identify--Outcomes

- ★ Identify potential digital content you may need to preserve
- ★ Treat the inventory as a management tool that grows as your program grows
- ★ Use it as a planning tool to prepare
  – e.g., staff, training, annual growth
- ★ Provides a basis for acquiring content, defining submission agreements, collections plans

Questions at this point

?

Resources

- ★ Society of American Archivists Jump In! Initiative: http://www2.archivists.org/groups/manuscript-repositories-section/jump-in-initiative
- ★ Short videos or 'elevator speeches' of the DPOE curriculum http://www.digitalpreservation.gov/education/curriculum.html
- ★ The Signal blog http://blogs.loc.gov/digitalpreservation/

Module 2

- ★ Select - *what portion* of that content will be preserved?

Why select content to preserve?

- Storage may be cheap, management is not … especially over time
- Quality of content
- Discovery and dissemination services … scale, scope, performance, sustainability
- Matching mission to content

Terms for Select

Different terms in different domains:

- Archives – appraisal and scheduling
- Libraries –selection
- Museums –acquisition

But there are common outcomes

Steps

1. Define and apply selection criteria
2. Document (and preserve) selection decisions
3. Implement your decisions

Define Selection Criteria

- ⋆ Acquisition or collection development policy
- ⋆ Departmental criteria (priorities, precedents)
- ⋆ Core record/content types (need no review)
- ⋆ Research criteria (interests, significance)
- ⋆ Uniqueness (only source)
- ⋆ Value (historical, evidential, can't reproduce)
- ⋆ Preserved elsewhere (avoid duplication)

Priorities

If you need to prioritize your review, consider:

- Most significant (producer, content)
- Most extensive
- Most requested
- Easiest (e.g., most familiar)
- Oldest (possible historical importance)
- Newest (possible immediate interest)
- Mandate (local, legislation, etc.)

Considerations during Review

Stop if or when the answer is 'no'…

1. Content
   - does the content have value?
   - does it fit your scope?
2. Technical
   - is it feasible for you to preserve the content?
3. Access
   - is it possible to make the content available, now or in the future?

Who would be on your dream team?

* ★ IT staff
* ★ Directors, other resource allocators
* ★ Content specialists--subject knowledge, records managers, curators, creators
* ★ Peers, colleagues
* ★ Experts (start thinking about who and where to find them. Your professional organization, nearby library school)
* ★ Student workers, volunteers

Documentation

Supplement inventory from Identify

* • Descriptions – more granular
  – Not item level, but enough to specify categories
* • Extent
  – How much content is there/will there be?
* • Use/access
  – When will content no longer be active?
* • Rights
  – Who owns rights to preserve and disseminate?

Select Module: Outcomes

Objectives:
* • How to gain control of possible content for planning
* • Where to start to begin to develop a sustainable program

Deliverables:
* • Expanded inventory of digital content
* • Agreements with producers: e.g., retention schedules, acquisition lists, submission agreements, deeds of gift
* • More team members/stakeholder buy-in

Questions?

?

Resources

* ★ Digital Preservation Management Tutorial. Based on OAIS, this tutorial focuses on overarching concepts. Also gives a good overview of OAIS and Trusted Digital Repositories. http://www.dpworkshop.org/dpm-eng/eng_index.html
* ★ Networking and Listservs:
  * ★ SAA listserv http://www.archivists.org/listservs/arch_listserv_terms.asp
  * ★ ALA Preservation and Reformatting Section http://www.ala.org/alcts/mgrps/pars

Today's Topics- 6 Modules

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?

Module 5

★ Manage - *what provisions* are needed for long-term management?

Why do we emphasize management?

Preserving Digital
Information (PDI),
1996

Commission on
Preservation and
Access & RLG

Balanced Management

An effective approach will address:

1. *Organizational* requirements and objectives
2. *Technological* opportunities and change
3. *Resources* – funding, staff, equipment, etc.

Organizational Issues: Skills

Skills that contribute to DP programs:

- Policy development
- Project management
- Repository/software management, programming
- Metadata management
- Legal expertise
- Marketing expertise

Organizational Issues: Policies

Benefits of a <u>preservation policy</u>:

- Specifies institutional commitment
- Developing policy builds DP team
- Demonstrates compliance – meet requirements
- Manages expectations – message to stakeholders
- Identifies issues and challenges
- Raises awareness
- Defines roles and responsibilities

Investing in Technology

★ <u>Prioritize:</u> weigh requirements to be met

★ <u>Assess:</u> define criteria to select appropriately

★ Sequence: identify steps to meet goals

★ Fund: decide when to own/join/share

★ <u>Anticipate</u>: look ahead, be prepared

★ Evaluate: measure outcomes and success

Adopting Technologies

Software should:

⋆ Work on widely-used platforms

⋆ Be well-supported by community and/or developers

Designated Funding

- Funds set aside for digital preservation
- Measurable indication of intent to preserve
- Challenging to do, but important
- Over time, contributes to track record
- May not be explicit (e.g., budget <u>line item</u>)
  … but must be able to make a compelling case

Succession Planning – willing to appoint an heir

DP Standards

Standards emerging since 1996 report :
- Trustworthy Digital Repositories, 2002
- Open Archival Information Systems (OAIS) Reference Model, 2003 and 2009 revision
- Preservation Metadata Implementation Strategies, 2005 plus updates (PREMIS)
- Trustworthy Repositories Audit and Certification (TRAC), 2011

Common practices are emerging and evolving

Trustworthy Digital Repository

A TDR should have these characteristics:
- community standards (OAIS Compliance)
- commitment (Administrative Responsibility)
- management (Organizational Viability)
- resources (Financial Sustainability)
- infrastructure (Technological Suitability)
- protection and control (System Security)
- documentation (Procedural Accountability)

Gap Analysis

Gap analysis as a management approach:
Where are we? (status)
Where do we want to be? (requirements)
Therefore, what gaps need to be filled? (objectives)

Outcomes

⋆What will standards conformance and good practice look like for your organization?

⋆How will your organization demonstrate good practice?

Module 6

★ Provide - *what considerations* are there for long-term access?

What is Long-term Access?
Preservation makes long-term access possible…

Preservation	Access
Relies on proven technologies to preserve digital objects across generations of technology	Relies on cutting edge technologies to provide best and fastest access at a point in time
Accumulates metadata over the life cycle to trace and preserve content	Selects metadata needed to use and understand content
Preservation systems create new versions of digital objects as needs and technology change over time	Access systems deliver the most appropriate version at any given time.
Purpose: ensure long-term access	Purpose: provide content to users
Focus: future users	Focus: current users

Requirements for providing content

Content should be delivered to users over time:

★ Easily – using current and known technologies

★ Coherently – well-documented and presented

★ Completely – intact and well-formed

★ Correctly – accurately representing deposits

★ Reliably – using well-managed technologies

★ Consistently – in accordance with policies

★ Fairly – with equity and precedent

Organizational Responsibilities

★ Develop and maintain comprehensive access policies

★ Manage preservation and access in parallel

★ Be transparent and compliant about access

★ Provide content to current and future users

★ Adapt new technologies for discovery and delivery

★ Manage legal issues throughout life cycle

Access Policies: Issues

★ Who is allowed to have access to content?

★ Are access policies equal for all content?

★ If not, how are categories managed?

★ How are exceptions/special requests handled?

★ How do users request/get access?

★ What options (if any) do users have?

Consider using FAQs as a step to develop policies

Access Policies: Implementation

★ Access decisions should be documented and translated into policy statements

★ Access policies should address requirements for preservation systems to produce access objects

★ Access policies should reflect and respond to new discovery/delivery issues that emerge

★ Preservation systems and procedures should incorporate and reflect access policy decisions

Roles for providing content

Roles that may contribute to long-term access include:

* Access service managers – define services
* Policy developers – capture access decisions
* Access system developers – enact decisions
* Access system managers – monitor and respond
* User support staff/systems – assist users
* System administrators – manage environment
* Users of content – understand and follow rules

Understand Users

* May be possible to track and respond to current users – e.g., usage, user surveys – who are your users?
* How do we anticipate needs of future users?
* User expectations are driven by delivery and discovery technologies they know and want
  - and we can't predict future technologies
* Preservation provides pathway from one generation of technology to the next
* How should digital content be packaged for delivery at specific points over time?

Sustainable Access

Effective and sustainable DP programs address:

* Value – understand and stress content value
* Roles – identify stakeholders and involve them
* Incentives – identify "carrots" for preserving

Identify and address costs across life cycle

Outcomes

* Clear access policies that address long-term access
* Links between preservation and access over time that ensure current access and long-term access
* Capacity to create dissemination packages from preservation packages using current technologies
* Awareness of and control over relevant rights management issues from creation and/or deposit on

What Next?

What are 3 things that you can do now to work on your institution's digital preservation program?

Questions?

?

Resources

* OAIS is extremely in-depth and the conceptual end can be disorienting. Brian Lavoie's introduction explains the model in clear language. www.dpconline.org/docs/**lavoie_OAIS**.pdf

* Digital Preservation Management Tutorial. Gives a good overview of OAIS and Trusted Digital Repositories. http://www.dpworkshop.org/dpm-eng/eng_index.html

* Trustworthy Digital Repositories https://www.oclc.org/resources/research/activities/trustedrep/repositories.pdf

* OAIS Reference Model (Open Archival Information System) http://public.ccsds.org/publications/archive/650x0m2.pdf

Thank you!

Lauren Goodley
lgoodley@txstate.edu

DPOE Baseline Principles (1-5)

1. Define the digital content within your scope of responsibility [Identify]

2. Specify the digital content you need/want to preserve [Select]

3. Establish requirements for storing files in preservation formats [Store]

4. Determine (and review) your best option for storing your content [Store]

5. Ensure that your content is secure during day-to-day activities [Protect]

DPOE Baseline Principles (6-10)

6. Work to ensure that your content is prepared for an emergency [Protect]

7. Develop (and review) plans for managing content over time [Manage]

8. Use policies to contain and develop your preservation program [Manage]

9. Remember that long-term access is the purpose of preservation [Provide]

10. Make sure the means to deliver content to users remains current [Provide]