An Introduction to Digital Preservation

Steps to Identify and Select Content

(First of 3 ASERL Webinars)
February 7, 2011

Jody L. DeRidder
University of Alabama Libraries
jlderidder@ua.edu

Content primarily from the Library of Congress DPOE Baseline Modules, version 2.0, Nov 2011
DPOE’s 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.”
DPOE 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?
Managing Content Over Time

DPOE Baseline Modules: Intro, version 2.0, Nov 2011
Why do we identify content?

**SCOPES!**

- Preservation requires an explicit commitment of resources
- Effective planning is based on knowing the extent of what will be preserved
- Identifying content is a first step to planning for current and future preservation needs
- An explicit inventory is the best way to identify content

Image from: http://en.wikipedia.org/wiki/Telescopic_sight
How will an inventory help?

Good preservation decisions are based on an *understanding* of the possible content to be preserved.

The Identify stage addresses:

“what content do I (or will I) have?”
Content Categories

Inventories should *include all relevant material*:

- Institutional records
- Special collections
- Scholarly content – licensed and open
- Research data
- Web content
- Digitized collections
Inventory Considerations

Inventory *content* is more important than style and format

- Inventor results should be:
  - *Documented*
  - *Usable*
  - *Available*
  - *Scalable*
  - *Current*
Inventory Tips

• Use available, familiar software to get started
  – *What software or tools do you already have?*
  – *What free or open source tools might be useful?*

• Be consistent, comprehensive, and concise
Level of Detail

- Inventories can be general to detailed
- Determine *appropriate level of detail for you*
- Factors:
  - Extent of content
  - Nature and location of content
  - Resources available
  - Deadlines

Photo courtesy of www.Ark-Of-Salvation.org
Format Types

An inventory should identify format types within categories of content - examples:

- Maps/geospatial
- Images
- Video
- Audio
- Text
- Drawings
- Web content
- Structured data

Indicate the range of file types when possible
Date Considerations

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
Location Is Important

- How do you specify online/offline location?
- Ability to change locations as content moves
- Location in storage system
- General location

*Be clear enough without going to extremes…*
Example entry:

- **Category:** Special Collections
- **Title/Description:** Railroad Photographs, SE U.S.
- **Type:** images, digitized
- **Format:** TIFF
- **Extent:** 242 GB; 2,250 images
- **Location:** archival server in Room A, Central IT
- **Coverage Dates:** early 1900’s
- **Creation date:** January-June 2006
- **Inventoried:** 12/15/2011, by Fred Jones
Outcomes of Identification

You will:

• 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
• Have a basis for acquiring content, making plans, and defining submission agreements
Questions for YOU!

• **What software** do you have at your disposal that you could use to create an inventory?

• **Where** would you store this inventory, and who would have access to it?

• **Who** at your institution needs to be involved in creating and updating this inventory?

• **When** in your institution’s normal workflow should capturing/updating this inventory occur?
Selecting Content for Preservation: Why do it?

- Storage may be cheap, but management is not especially over time.
- Sustaining the quality of content takes effort.
- Continually changing discovery and dissemination services will be needed as hardware and software change... think scale, scope, performance, sustainability.
Different terms used in different domains:

- Archives – appraisal and scheduling
- Libraries – e.g., selection
- Museums – e.g., acquisition

*But all have common outcomes*
Steps

1. **Review** your potential digital content
2. **Define and apply** selection criteria
3. **Document** (and preserve) selection decisions
4. **Implement** your decisions

...think of this as weeding your garden
Selection Criteria:
matching mission to content...

- Acquisition or collection development policy
- Departmental criteria (priorities, precedents)
- Research criteria (interests, significance)
- Uniqueness (only source)
- Value (historical, evidential, can’t reproduce)
Practical Considerations

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?
Selection starts at the beginning…

*Prevent later headaches!*

Contact content creators
(as needed)

- Arrange a convenient time for them
- Prepare brief statement of outcomes
- Identify list of materials to review with them
- Send a reminder before the meeting
- Document the results and send them a copy
Documenting

Add more information to your inventory!

• **Descriptions** – more granular
  – Not item level, but enough to specify categories

• **Extent**
  – How much content is there/will there be?

• **Use**
  – When will content no longer be active?

• **Rights**
  – Who owns rights to preserve and disseminate?
Priorities

If you need to prioritize your review, consider:

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

May benefit from a team-based approach!!!

• Analytical skills
  – Review and understand content
  – Determine relationships and significance
  – Make sound and consistent judgments

• Interpersonal skills
  – Communicate clearly and compellingly

• Technical Skills
  – Determine feasibility of preservation and access
Outcomes of Selection

Objective:

• Gain control of possible content for planning
• Develop a sustainable program

Possible products of selecting content:

• More detailed information in inventories of content to preserve
• Agreements with producers: e.g., retention schedules, acquisition lists, submission agreements
Questions for YOU!

• **Identify and rank the priorities** you think your institution should use in selecting which content to preserve.

• **Who** at your institution needs to be involved in setting those priorities and carrying out this procedure?

• **How** would implementation be integrated into existing work flows?
An Example… in Special Collections

FIRST: Who needs to be involved, and how?

**Engage:**
- Management
- Archivists
- Collection development
- Technical support

-- explain the issues –
*Relate it to your institution’s mission,*
*meeting researcher needs…*
*(and the donor agreements?)*
An Example… in Special Collections

SECOND: Procedures, Policies

- Draw up preliminary form of spreadsheet
- Identify shared location where it will be kept
- Identify who should add content when as part of their workflow
- Identify who should modify the inventory when, as part of their workflow
An Example... in Special Collections

THIRD: The Nitty Gritty: do this in stages!

1. Interview archivists for known locations
2. Review finding aids to identify general locations
3. Hunt through boxes to locate specific media
4. Find the hardware/software to open the media
5. Open the media to locate and identify all the digital files
An Example… in Special Collections

The Nitty Gritty gets really gritty

- Change folder settings and view details… you’ll need them.
- Start at the top, and work down one directory at a time, in order.
- List contents of each directory
Selection Criteria: matching mission to content...

- Acquisition or collection development policy
- Departmental criteria (priorities, precedents)
- Research criteria (interests, significance)
- Uniqueness (only source)
- Value (historical, evidential, can’t reproduce)

Identification & Selection ideally takes place WITH the DONOR or creator BEFORE acceptance!!!!
Practical Considerations

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?

Ask the curator!

Ask the IT person!
Steps

1. **Review** your potential digital content
2. **Define and apply** selection criteria
3. **Document** (and preserve) selection decisions
4. **Implement** your decisions

...think of this as weeding your garden
What about born-digital content?

- Students and faculty: papers, presentations
- Staff: university or government documents for archives
- Faculty creating data sets during research

Content creators NEED GUIDANCE to create digital documents that are preservation-ready.
Why are we here, after all?

S. R. Ranganthan (1931), paraphrased:

- **Information is for use.**
- Every user his / her information.
- Every piece of information its user.
- Save the time of the user.
- **The library is a growing organism.**
DPOE Baseline Principles (1-2)

• **Session #1: Identify & Select**

1. Define the digital content within your scope of responsibility **[Identify]**
2. Specify the digital content you need/want to preserve **[Select]**
DPOE Baseline Principles (3-6)

3. Establish requirements for storing files in preservation formats [Store]
3. Determine (and review) your best option for storing your content [Store]
3. Ensure that your content is secure during day-to-day activities [Protect]
3. Work to ensure that your content is prepared for an emergency [Protect]

Session #2: Considerations for Storage & Protection of Content

February 14, 2012 | 11am ET
https://www2.gotomeeting.com/register/461694690
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]
More Resources

“Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems”

- Online Tutorial:
  http://www.dpworkshop.org/dpm-eng/eng_index.html
- Survey of Institutional Readiness:

**Exercises and the DPOE Resource lists:**
http://jodyderidder.com/resources/

EVALUATE THIS WEBINAR!
http://www.surveymonkey.com/s/BasicDigiPres1