FROM THE DIGITAL DARK AGES TO A DIGITAL RENAISSANCE

THE ART OF SELECTING DIGITAL CONTENT TO PRESERVE

Part I. Identify
MANAGING DIGITAL CONTENT OVER TIME

Identify - what digital content do you have?
Select - what portion of that content will be preserved?
Store - how should your content be stored for the long term?
Protect - what steps are needed to protect your digital content?
Manage - what provisions are needed for long-term management?
Provide - how should your content be made available over time?
WHY DO WE IDENTIFY CONTENT?

- Identifying content is a first step necessary to inform decisions about preservation.
- Digital preservation requires an explicit commitment of resources.
- Not all digital content in and around an organization will be preserved.
Good preservation decisions are based on an *understanding* of possible content

**Identify** addresses:
“what content do I have?”
QUESTION

Before you start - think about your organization –

Who can help? Whose support do you need?
INVENTORY CONSIDERATIONS

- Inventory **content** is more important than style and format

- Inventory results should be:
  - Documented: inventory should *actually exist*
  - Usable: simple format for sort, list, etc.
  - Available: accessible where needed
  - Scalable: make choices in Select process
  - Current: update periodically
INVENTORY TIPS

- Use available, familiar software to get started
  - What software or tools do you already have?
  - What free tools may also be useful?
- Be consistent, comprehensive, and concise
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?
HOW MUCH DETAIL TO INCLUDE

- Inventories can be general to detailed

- Choose the appropriate level of detail

- Factors in determining level of detail:
  - Extent of content to be inventoried
  - Nature and location of content
  - Resources available to complete inventory
  - Timeframe, deadlines for completion
An inventory should identify format types within categories of content

- Images
- Video
- Audio
- Text
- Structured data

- Maps/geospatial
- Drawings
- Web content
Inventories should note:

- Date of inventory – and updates
- Date of files – created or modified
- Dates covered in content – even approximate
- Date received – if relevant, possible
LOCATIONS

Locations are important – consider:

- Method to specify online/offline location
- General location – e.g., with us, with creator
- Ability to change locations as content moves
- Method storage systems use to note 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
IDENTIFY OUTCOMES

- Identify potential digital content the organization may need to preserve
- Management tool for digital content over life cycle – supplemented in stages
- Planning tool to prepare for relevant digital content – staff, training, annual growth
- Basis for acquiring content, submission agreements, development plans
Does your institution currently have a written, documented inventory of its digital content/holdings?

- Yes
- No
- Hmmmm….I don’t know!
GET STARTED! A FEW QUESTIONS TO START WITH…

- 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?
IDENTIFY RESOURCES

Digital Content Inventories, Examples:

- California Digital Library (CDL) / UC Libraries Digital Assets Submission Inventory
  www.cdlib.org/services/dsc/contribute/docs/submission.inventory.rtf

- University of Kansas Data Collection Manual for Digital Assets
  http://kuscholarworks.ku.edu/dspace/bitstream/1808/166/6/Appendix%20E%20-%20Data%20collection%20manu%20draft%202.pdf

File Formats & Identification:

- "File Format Identification and Validation Tools," Roy Lechich, Yale University, February 2007,
  http://www.imaging.org/ist/publications/reporter/articles/Rep22_4_Arch2007_CHOU.pdf

- Format Identification, Validation, Characterization and Transformation in DAITSS, Carol Chou:


- DROID, PRONOM and Preservation Planning: Anyone can do it!, by Alison Heatherington, 2009:
  http://www.dpconline.org/docs/events/0906roadshowYorkHeatherington.pdf

- PRONOM - A Practical Online Compendium of File Formats, by Jeffrey Darlington, RLG DigiNews, 2003:

- Sustainability of Digital Formats, Planning for Library of Congress Collections:
  http://www.digitalpreservation.gov/formats/sustain/sustain.shtml
Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems (Online Tutorial): http://www.dpworkshop.org/dpm-eng/eng_index.html

FROM THE DIGITAL DARK AGES TO A DIGITAL RENAISSANCE

THE ART OF SELECTING DIGITAL CONTENT TO PRESERVE

Part II. Select
Identify - what digital content do you have?

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

Store - how should your content be stored for the long term?

Protect - what steps are needed to protect your digital content?

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

Provide - how should your content be made available over time?
WHY SELECT CONTENT TO PRESERVE?
GETTING ON THE SAME PAGE

Different terms in different communities:

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

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
STEPS

1. *Review* inventory and establish priorities

2. *Define and apply* selection criteria

3. *Document* (and preserve) selection decisions

4. *Implement* your decisions
Here are some things to consider when prioritizing your inventory:

- Most significant (producer, content)
- Most extensive
- Most requested
- Easiest (e.g., most familiar)
- Oldest (possible historical importance)
- Newest (possible immediate interest)
- Mandated by law
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)
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?
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
Enrich inventory information:

- **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?
Possible products of selecting content:

- Expanded inventory that might include retention schedules, acquisition lists, submission agreements

Objective:

- Gain control of your content
- Develop a sustainable program
QUICK POLL!

Does your institution currently have a written, documented selection/acquisition policies for digital content/digital collections?

- Yes
- No
- Hmmmm….I don’t know!
GET STARTED! A FEW QUESTIONS TO START WITH…

- 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 workflows?
ADDITIONAL RESOURCES

- Copyright and Cultural Institutions [http://tinyurl.com/3w7gnqg](http://tinyurl.com/3w7gnqg)
- Copyright Crash Course [http://copyright.lib.utexas.edu/](http://copyright.lib.utexas.edu/)
- Digital Copyright Slider [http://librarycopyright.net/digitalslider/](http://librarycopyright.net/digitalslider/)
Digital Preservation Outreach & Education

SELECT RESOURCES

Digital Resources & Selection


- Data Asset Framework http://www.data-audit.eu/

Records Appraisal


Assessment Criteria for Digitization

- Oxford University, 1999: http://www.bodley.ox.ac.uk/scoping/assessment.html
### SECTION 1: PROGRAM OVERVIEW

| A. Introductory description of program or collection(s) |
| B. Organization or program mission statement |
| C. User groups |
| D. Team |

**MODULE: MANAGE**

Include a brief description of the program/collection, a description of how the program/collection serves the institution’s larger mission, identification of user groups, and names/contact information for team members and collaborators responsible for the collection.

### SECTION 2: SELECTION, ACQUISITION & MAINTENANCE

| A. Scope of collection(s) |
| B. Selection process |
| C. Content types & formats |
| D. Deselection guidelines & processes |
| E. Rights management |

**MODULES: IDENTIFY & SELECT**

Briefly describe the scope of the collection (can be as broad or specific as needed), the established selection criteria and selection process (by committee, institutional mandate, etc.), any guidelines or processes for deselection if applicable (e.g., retention schedules), and rights management/legal restrictions impacting materials in the collection.

### SECTION 3: METADATA

| A. Descriptive metadata |
| B. Administrative metadata |
| C. Preservation metadata |

**MODULES: STORE & PROTECT**

Metadata is beyond the scope of this 3-hour workshop. However, metadata schemas and standards used to describe, manage, and preserve collections are important to address in planning documentation.

### SECTION 4: PRESERVATION

| A. Description of preservation system/repository/strategy |
| B. Succession plan |
| C. Disaster plan |

**MODULE: PROVIDE**

Describe your preservation repository or strategy – multiple copies in multiple locations, network membership, cloud services, vendor-provided repository, open-source/in-house system, etc. Discuss preferred file formats for preservation and file validation/security measures. Succession planning and disaster planning efforts may be addressed here as well.

### SECTION 5: PRESENTATION AND ACCESS

| A. Access policies (by collection/user group) |
| B. Access system |
| C. Discovery |

**MODULE: PROVIDE**

Document access policies for content and user types. Describe your access system – how do users get to your content? Is your content accessible online? Or is it inaccessible and stored within a dark archive? Are there legal restrictions impacting the accessibility of certain materials? How are users able to find, search, and browse the collections?

### SECTION 6: PLANNING AND ASSESSMENT

| A. Strategic planning schedule |
| B. Policy review schedule |
| C. Inventory schedule |
| D. Program assessment parameters & schedule |

**MODULE: MANAGE**

Your plan is a living document that will grow and change over time. Determine and document timeframes for strategic planning, reviewing policies, and conducting inventories (once every 2 years? 5 years? 10 years?). Outline basic parameters for assessing your program on a regular basis, as well as any auditing tools or services that might be utilized in program evaluation.

### APPENDICES

Appendices may include service/collaboration agreements, an organization chart, inventory/selection documents, rights documentation/submission agreements, or system documentation for your preservation/access repositories.
Keep in touch!

And for those viewing the recorded version of this webinar on the ALCTS website, **we want to hear from you!** E-mail us to request the resource lists and ask questions.

- Brenda Miller, Hartford History Center, bmiller@hplct.org
- Sarah Rhodes, National Geographic Library, srhodes@ngs.org

Images courtesy of Microsoft Office Free Clip Art and Yahoo’s flickr Creative Commons.

A special thanks to DPOE team member Jody L. DeRidder for her assistance in elucidating the Identify and Select DPOE modules.