Digital Preservation: Principles and Practice

Planning and Building a Digital Collections Program
NERCOMP, May 2012

bit.ly/nercomp_digitalcollect
Background: 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.”

http://www.digitalpreservation.gov/education/
DPOE Model

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
IDENTIFY

PRINCIPLE 1:
DEFINE THE DIGITAL CONTENT WITHIN YOUR SCOPE OF RESPONSIBILITY
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
- 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** 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 during **Select**
  - Current: update periodically
SELECT

PRINCIPLE 2:
SPECIFY THE DIGITAL CONTENT YOU NEED/WANT TO PRESERVE
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
Steps

• Review your potential digital content
• Define and apply selection criteria
• Document (and preserve) selection decisions
• Implement your decisions
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)
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?
STORE

PRINCIPLE 3:
ESTABLISH REQUIREMENTS FOR
STORING FILES IN PRESERVATION
FORMATS
What are storage needs?

Archival Storage manages content as objects

Digital content (files + metadata = object)

• May include any type of content
  – e.g., images, text, sound, video, maps

• Requires some identification and description
  – Captured as metadata

• Needs at least two copies at least two places
Well-managed Collections

Well-managed status makes preservation easier

Sample characteristics of well-managed:

• Basic information about each deposit
• Minimal metadata for objects (you define)
• Common (or normalized) file formats
• Controlled and known storage of content
• Multiple copies in at least 2 locations
STORE

PRINCIPLE 4:
DETERMINE (AND REVIEW) YOUR BEST OPTION FOR STORING YOUR CONTENT
Storage Media Options

• Content (objects) are kept on storage media
• Options include: online, near-line, offline
• Factors for choosing options include
  – Cost (available resources for preservation)
  – Quantity (size and number of files)
  – Expertise (skills required to manage)
  – Partners (achieving geographic distribution)
  – Services (outsourcing)
Storage Considerations

- Multiple, geographically distributed copies
- Storage Partners
- Hosted services, e.g.

This is a service to make it easy for organizations to use cloud services to manage content over time.
Objectives

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
• Monitor copies of content for errors/change
• Plan for media replacement
PROTECT

PRINCIPLE 5:
ENSURE THAT YOUR CONTENT IS SECURE DURING DAY-TO-DAY ACTIVITIES
What are we protecting content from?

- Change and loss – accidental and intentional
- Obsolescence – as technology evolves
- Inappropriate access – e.g., confidential data
- Non-compliance – standards and requirements
- Disasters – emergencies of all kinds
Everyday Protection

• Know where your content is located
  – Onsite and offsite; online and offline
• Know who can have access to it
  – DP staff, IT staff, others?
• Manage authentication information
  – For staff, depositors, users
• Track and review usage then adjust practices
  – Web use, internal use and activities, maintenance
PROTECT

PRINCIPLE 6:
WORK TO ENSURE THAT YOUR CONTENT IS PREPARED FOR AN EMERGENCY
Readiness

Proper planning should allow you to:

• Prevent – undesirable outcomes
• Predict – most likely risks and threats
• Detect – errors, problems, damage
• Respond – with appropriate measures
• Repair – damage or possible loss
Emergency Protection

• Engage in ongoing disaster planning
  – Establish committee and share information
  – Develop and maintain documents

• Identify possible outcomes and prepare
  – e.g., server goes down, media is damaged
MANAGE

PRINCIPLE 7:
DEVELOP (AND REVIEW) PLANS FOR MANAGING CONTENT OVER TIME
Why do we emphasize management?

Preserving Digital Information (PDI), 1996

Commission on Preservation and Access & RLG
Balanced Management

An effective approach will address:

• Organizational requirements and objectives
• Technological opportunities and change
• Resources – funding, staff, equipment, etc.
MANAGE

PRINCIPLE 8:
USE POLICIES TO CONTAIN AND DEVELOP YOUR PRESERVATION PROGRAM
Trusted 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)
Policies

Benefits of a preservation policy:

• 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
Planning

- Preservation Planning (ongoing)
- Self-assessment (internal process)
- Audit (external review by peers)

Also

- Business Continuity (Protect)
- Disaster Planning (Protect)
PROVIDE

PRINCIPLE 9:
REMEMBER THAT LONG-TERM ACCESS IS THE PURPOSE OF PRESERVATION
What is Long-term Access?

Preservation makes long-term access possible...

Preservation

- proven
- accumulate
- access over time
- future users

vs.

- technologies
- metadata
- purpose
- focus

Access

cutting edge
relevant now
access now
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
PROVIDE

PRINCIPLE 10:
MAKE SURE THE MEANS TO DELIVER CONTENT TO USERS REMAINS CURRENT
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

Ongoing Access

• Clear access policies that address long-term access
• Links between preservation and access over time that 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
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]
Next Steps

• Think about your current situation
• Write down the top 2 or 3 things you think you most need to do – not more than 3
• For each one, think through:
  – Who you might work with at your organization or with a colleague/friend somewhere else
  – Challenges you might face and ways to address those
  – How much time you might need (available time)
  – What the specific outcome will be
• Pick the one to start with
THANK YOU!

bit.ly/nercomp_digitalcollect