

Digital Preservation: What Are the Critical Issues?

Terry Eastwood

Archival Studies

The University of British Columbia

My Research

- 1994-97, project on management of digital records by their creating body
- 1998-2001, InterPARES 1 on long-term preservation of traditional organizational records in digital environment
- 2002-06, InterPARES 2 on reliability, authenticity, and accuracy of dynamic, interactive and experiential records in artistic, scientific, and government realms

The Problem

- In the digital environment, there are no originals, only copies
- Technological obsolescence and media fragility can render it impossible to reproduce copies
- The means of overcoming obsolescence can impair the reliability, authenticity, and accuracy of digital material

Some Conventional Wisdom

- Preservation starts at creation, and
- Must address appraisal at the outset with the creator's involvement
- Metadata describing context, structure, and content is an essential component of preservation
- Material in standard format, or transformable into such, is a godsend

What I am going to do

- Explain in simplified terms what we have discovered in producing a life cycle model of all the activities undertaken in managing the chain of preservation of digital records
- Explain very generally our conception of the classes of metadata associated with creation, maintenance, and long-term preservation
- Draw some conclusions about the critical issues in digital preservation

Difficulties I face

- Our work has been on records, i.e., documents made or received in the conduct of affairs, and preserved.
- Publications, Web sites, datasets, and databases are often not records, so much of our thinking about the life cycle may not be relevant to such materials.
- Nevertheless, we'll see what we can learn

What is a life cycle?

- Usually based on the notion of continuous management of digital resources from the time of their conception and creation through all other phases of their existence, including the long-term preservation phase to ensure their reliability, authenticity, accuracy, accessibility, and intelligibility.

What are these phases?

- In the library world:
- “creation, acquisition, cataloguing/identification, storage, preservation and access.” Hodge 1999
- The Library of Congress: “refers ...to the progressive technology and workflow requirements needed to ensure long-term sustainability of and accessibility of digital objects and/or metadata.”

Traditional Records Life Cycle

- Active phase: making and receiving records, keeping them in record keeping systems, maintaining and using them for current business purposes
- Semi-active phase: maintaining them for continued reference in low cost storage
- Inactive phase: transfer to an archival authority, followed by identification, description, preservation, and access for research

InterPARES Approach

- To characterize all of the activities involved in life cycle management of digital records
- Perspective is on the activities that effect the preservation of the reliability, authenticity, and accuracy of digital records

Four Main Activities

- Set up a framework of policies, strategies, procedures, and instruments to manage the chain of preservation (for both creator and preserver)
- Manage creation activities
- Manage maintenance activities
- Manage long-term preservation activities

Management Framework

- Attuned to the kind of creators and digital documents or data they are producing
- Policies etc. to ensure that the creator institutes necessary preservation measures when (1) making digital entities and (2) maintaining digital entities
- Policies etc. to ensure that the preserver institutes necessary measures for long-term preservation

Creating Digital Entities

- Some things to guide creators to do:
 - Adopt, whenever possible, hardware, software, data standards, and system administration procedures that facilitate preservation
 - Adopt metadata schemes that characterize aspects of data context, structure, and content for preservation purposes when they would not otherwise be evident

Maintaining Digital Entities

- Some things to guide creators to do:
 - Manage storage of digital entities to ensure they are not intentionally or unintentionally altered
 - Document actions taken to update records (future users base assumption of authenticity and accuracy on knowledge of such updating actions)
 - Keep so as to be able to transfer relevant documentation to the preserver

Preserving Digital Entities

- The preserver carries out these activities:
 - Setting a preservation management framework
 - Appraising digital entities for permanent preservation
 - Acquiring digital entities
 - Preserving accessioned digital entities
 - Reproducing digital entities for users

Preservation Management Framework

- Some things preservers need to do:
 - Establish acquisition policy and strategy, and appraisal criteria
 - Determine procedures to monitor creator's production of digital entities
 - Establish preservation technological requirements, strategies, and procedures
 - Establish a descriptive/retrieval system

Appraising Digital Entities

- Analyzing information about digital entities
- Assessing their value*
- Determining the feasibility of preservation*
- Making an appraisal decision*
- Monitoring appraised digital entities when they have been identified for preservation

* = more to come on these

Assessing Value

- Involves making a decision by:
- Assessing continuing value
 - According to set policy, strategy, and criteria
 - Taking into account creator's needs
 - Taking into account societal needs
- Assessing reliability, authenticity, and accuracy
 - Assessing the grounds for assuming that the entities can be trusted to be what they purport to be, have not been corrupted, and are accurate as to content

Determining Feasibility of Preservation

- Vital and unique aspect of the process
- Involves analyzing the characteristics of the documents or data that must be preserved to maintain reliability, authenticity, and/or accuracy to determine preservation requirements
- Then, reconciling these preservation requirements with the current and expected preservation capabilities

Acquiring Digital Entities

- To acquire digital entities requires that the preserver:
 - Verifies who transferred them
 - Verifies that the transfer is correct and complete, and
 - Confirms the feasibility of preservation
- Then and only then does the preserver take the material into custody

Preserving Digital Entities

- Preserving documents or data involves:
 - Managing preservation information about those entities
 - Managing storage of those entities
 - Taking the necessary actions to ensure long-term preservation

Managing Preservation Information

- The preserver builds a system that:
 - Compiles information for preservation, description, and output
 - Describes acquired entities
 - Keeps information about preservation actions (e.g. migration, reformatting or refreshing storage media, etc.)
 - Provides information about preserved digital entities and their technological character

Managing Storage of Digital Entities

- Entities might be made up of digital components, each component being something that can be stored and retrieved and assembled with other components to reproduce the entity
- Storage has to be monitored & storage problems corrected
- Technological change in the preservers environment may cause the need to migrate materials or reformat or refresh storage of digital components on storage media

Reproducing Digital Entities

- According to the preservation strategy determined at the time of feasibility analysis
- May involve either bringing the components of the entities together for the user and presenting them as copies or packaging the components together with the information necessary for the user to reproduce them

IP2 Picture of Life Cycle Metadata

- The next slide presents a diagram that depicts the various classes and sub-classes of metadata that could be generated during the life cycle management of digital materials, much of which serves to provide a guarantee the reliability, authenticity, accuracy, and intelligibility of those materials

**Record Creation
A2**

**Record Keeping
A3**

Record Preservation A4

Identity Metadata

**Description
Metadata**

Integrity Metadata

**Retrieval
Metadata**

**Creation
Process
Metadata
(outgoing,
transmission,
receipt)**

**Protective
Procedure
s
Metadata**

**Feasibility
Confirmation
Metadata**

**Presentation
Metadata**

**Access
Metadata**

**Transfer
Authorization
Metadata**

**Package
Output
Metadata**

**Destructio
n Metadata**

Transfer Metadata

**Record
Accession
Metadata**

**Transfer
Verification
Metadata**

**Preservatio
n Action
Metadata**

Critical Issues

- Work with creators of digital resources
- Develop detailed and rigorous appraisal standards to ensure resources are not wasted preserving materials of dubious value
- Record and preserve information about the context, structure, and content of digital entities to enable future users to understand and trust them

To View InterPARES Model

- The activity model consists of many diagrams done in IDEF-0 and a glossary defining all terms used in all the diagrams

- To view it go to:

http://www.interpares.org/ip2/ip2_models.cfm#