

## Introduction to Digital Preservation: Issues, Principles, Players, & OAIS

Howard Besser  
<http://besser.tsoa.nyu.edu/howard/>  
<http://www.nyu.edu/tisch/preservation/>

Dig Preserv class intro, 9/9/09 1

## How long we keep things

- Companies keep information for **days, or even years**
- Individuals keep things for **years, or a lifetime**
- Archives, Libraries, and museums keep things for **hundreds of years**

*Cultural Institutions have a much greater responsibility for preservation!*

Dig Preserv class intro, 9/9/09 2

## But who is preserving today's "born-digital" works?

- **In the past, we knew about history by finding written documents:**
  - Changes between different drafts of a scientific or literary paper
  - Letters and correspondence between a scientist (or literary figure) and colleagues (that both helps contextualize the work, and lets us see changes in thought processes or discovery)
- **But today, these documents are not on paper! They are in the form of:**
  - Email correspondence
  - Word processing files that do not show changes between drafts/versions
- **Who will take responsibility to save these works for future study?**

Dig Preserv class intro, 9/9/09 3

## Won't Talk about Digital Restoration (VidiPax)



Dig Preserv class intro, 9/9/09 4

## Paradigms Shifts needed

|                              | <u>Old</u>               | <u>New</u>                                    |
|------------------------------|--------------------------|---|
| <b>Physical preservation</b> | atmospheric cntrl        | ongoing mgmt                                  |
| <b>What to save?</b>         | artifact                 | idea + ancillary material & documentation     |
| <b>Cataloging</b>            | Individual work in hand  | FRBR  |
| <b>Later access</b>          | Artifact & documentation | Restaging, ancillary material & documentation |

Dig Preserv class intro, 9/9/09 5

## Preserving Difficult Electronic Works: Moving Images, Websites, and Electronic Art-

- How are new works even more problematic than older forms of moving image material?
- Issues with Digital Preservation
- Issues with New Works
- Technical & Conceptual Approaches to solutions
- Efforts to watch (projects, standards)
- **Paradigm shifts needed**

Dig Preserv class intro, 9/9/09 6

## Conventional Works

- Manuscripts, books, paintings, sculpture
- We have a good sense of what the original object is
- Objective is to make object itself endure (temperature /humidity control, chemicals/pigments/fibers/adhesives, ...)
- Goal is to keep object as close as possible to original state (though occasionally controversy arises over whether to let aging show)

Dig Preserv class intro, 9/9/09

7

## Electronic Media

- Video, audio, digital, new media
- Often difficult to determine what the original object is
- Difficult to make the original object endure (magnetic particle deterioration, warping, etc.)
- Even if we could make the original object endure, we wouldn't have the infrastructure to view it in the future
- Need to develop a paradigm shift from preserving the original object to preserving info content
- Need to pay more attention to maintaining authenticity and replicating user experience

Dig Preserv class intro, 9/9/09

8

## Electronic Art in general is not like canvas paintings

- May include
  - Moving image materials
  - Multimedia
  - Interactive programs (including hypertext novels & games)
  - Computer generated art
- Most electronic art works share some common characteristics with other "strange" works like
  - Performance Art
  - Conceptual Art
  - Site-specific installations
  - Experiential Art

Dig Preserv class intro, 9/9/09

9

## The Short Life of Digital Info: Digital Longevity Problems-

- ♣ Disappearing Information
- ♣ The Viewing Problem
- ♣ The Scrambling Problem
- ♣ The Inter-relation Problem
- ♣ The Custodial Problem
- ♣ The Translation Problem

Dig Preserv class intro, 9/9/09

10

## The Viewing Problem

- ♣ Digital Info requires a whole infrastructure to view it
- ♣ Each piece of that infrastructure is changing at an incredibly rapid rate
- ♣ How can we ever hope to deal with all the permutations and combinations

Dig Preserv class intro, 9/9/09

11

## Viewing Problem

- Requires new file formats and new physical strata at regular intervals
- Needs a serious Managed Environment
- Main InterPARES finding--the need for complete lifecycle management
  - archivist needs to be involved when record is created and throughout active life

Dig Preserv class intro, 9/9/09

12

## The Scrambling Problem

Dangers from:

- ✿ Compression to ease storage & delivery
- ✿ Container Architecture to enhance digital commerce

Dig Preserv class intro, 9/9/09

13

## The Inter-relation Problem

- ✿ -Info is increasingly inter-related to other info
- ✿ -How do we make our own Info persist when it points to and integrates with Info owned by others?
- ✿ -What is the boundary of a set of information (or even of a digital object)?

Dig Preserv class intro, 9/9/09

14

## The Custodial Problem

- ✿ In the past, much of survival was due to redundancy
- ✿ How do we decide what to save?
- ✿ Who should save it?
  - ✿ Mellon-funded E-Journal Archives
- ✿ How should they save it?-

Dig Preserv class intro, 9/9/09

15

## The Custodial Problem:

How to save information?

- ✿ Methods for later access
  - ✿ Refreshing
  - ✿ Migration
  - ✿ Emulation
- ✿ Issues of authenticity and evidence

Dig Preserv class intro, 9/9/09

16

## The Translation Problem

- ✿ Content translated into new delivery devices changes meaning
  - A photo vs. a painting
  - If Info is produced originally in digital form in one encoded format, will it be the same when translated into another format?
  - Behaviors

Dig Preserv class intro, 9/9/09

17

The Translation Problem

## Thinking of the Future (1/2)

- Screens will be different resolutions and different aspect ratios
- CRTs won't exist
- A decade or 2 from now, today's user interfaces will look like arrow-key navigation looks like today

Dig Preserv class intro, 9/9/09

18

The Translation Problem

## Thinking of the Future (2/2)

- Today's streaming media are small windows, slow speeds
- As bandwidth increases, viewers will expect higher quality streams
- Creators may need to consider how they'll be able to deliver higher-bandwidth streams
  - Delivery Derivatives vs. Masters encoded w/standards
  - May also want to re-edit the piece to take advantage of changes in technology, viewer expectations, society-

Dig Preserv class intro, 9/9/09

19

## Screen Formats



## Responding to serious Longevity Problems

- ❖ Previous formats required little ongoing intervention (remote storage facilities, Iron Mtn); digital formats require intense ongoing management
- ❖ Need for:
  - ❖ Preservation Repositories
  - ❖ Preservation Metadata

Dig Preserv class intro, 9/9/09

21

## Issues with new works-

- What is the work?
- Complexity of rich media
- Difficulty of making the work last

Dig Preserv class intro, 9/9/09

22

## Complexity of Rich Media

- Works often have artistic nature (including video games)
- Enormous number of elements can, at times, be very important to preserve (pacing, original artifact, elements used to construct the artifact)
- Too complex to save every one of these aspects for every type of material
- Importance of saving documentation

Dig Preserv class intro, 9/9/09

23

## Special Characteristics of Electronic Works

- What Really is the Work?
- Disappearing software
- Enormous number of elements can, at times, be very important to preserve (randomness, interactivity, pacing, color, format, original artifact, elements used to construct the artifact)
- Pieces and Boundaries
- Recontextualization (Postmodernism)--which rendition to save?
- Dynamic & Lack of Fixity (evolving works)
- Interactivity
- Historical context
- Difficulty of authentication over time

Dig Preserv class intro, 9/9/09

24

## Documentation & Preservation: What are we trying to do?

- Show the work the way people saw and interacted with it when it was first created

(may be impossible; in the past, the artifact and how one interacted with it didn't change much, so preservation and documentation were relatively straightforward)

- Show documentation of the work and people interacting with it when it was first created
- Reinstall/Recreate/Reinact the work

Dig Preserv class intro, 9/9/09

25

## Technical & Conceptual Approaches to Solutions-

- Save the Hardware & Software
- Emulate
- Migrate
- FRBR
- Artist Intentions

Dig Preserv class intro, 9/9/09

26

## Save the Hardware & Software-

- A huge undertaking
- Computer Museum
- Broderbund

Dig Preserv class intro, 9/9/09

27

## Old Video Formats



Dig Pres

28

## Old Digital Formats



Dig Pres

29

## Save the Hardware & Software

- A huge undertaking
- Computer Museum
- Broderbund

Dig Preserv class intro, 9/9/09

30

## Possible endless need for reformatting implies

- Possible loss with each generation
- Requires managed environment

Dig Preserv class intro, 9/9/09

31

## Managed Environment

- More than temperature & humidity control
- Periodic monitoring of the works
- Periodic monitoring of the technical environment for viewing the works (software, systems, hardware)
- Trusted repositories

Dig Preserv class intro, 9/9/09

32

## Approaches to Solutions-

- ~~Save the Hardware & Software~~
- Emulate
- Migrate

Dig Preserv class intro, 9/9/09

33

## Conceptual Approaches to Digital Preservation

- **Refreshing** always necessary due to volatility of physical strata
  - Impact on evidential value
- **Migration** -- advantages & disadvantages
- **Emulation** -- advantages & disadvantages
- And will need a long-term managed environment

Dig Preserv class intro, 9/9/09

34

## Migration

- Wordstar to Word 1 to Word 3, ...
- -Tables and complex features often get corrupted
- -Need to repeat every 4-5 years (maybe forever)
- +We know how to do this ourselves
- +If there's a problem, we can catch it soon

Dig Preserv class intro, 9/9/09

35

## Emulation

- Keep the Wordstar file format, but write emulators to make it work in newer environments
- +A better chance of carrying over complexity
- +Many more features can survive
- -Problems may not be caught until it's too late
- -Specialists and a whole infrastructure of emulators required
- -Serious © problems (reverse engineering?)

Dig Preserv class intro, 9/9/09

36

## Managed Environment

- More than temperature & humidity control
- Periodic monitoring of the works
- Periodic monitoring of the technical environment for viewing the works (software, systems, hardware)
- Trusted repositories

Dig Preserv class intro, 9/9/09

37

## Incorporate parts of Functional Requirements for Bibliographic Records (FRBR)

- work
- expression
- manifestation
- item

Dig Preserv class intro, 9/9/09

38

## Standards for encoding artists intentions

(group efforts w/i Cult Heritage community)

- Media Matters--SFMOMA/MoMA/Tate
- Variable Media Initiative
- Seeing Double Exhibition, & Symposium
- IMAP
- Artists Interviews Project, Netherlands Institute for Cultural Heritage 1998-1999, **Modern Art: Who Cares** (<http://www.icn.nl/english/6.4.2.html>)
- TechArcheology: A Symposium on Installation Preservation (SFMOMA)

Dig Preserv class intro, 9/9/09

39

## Multiple Approaches

- Capture the documentation
- Freeze a final version (eg. PDF of CAD)
- Capture the full functionality

### Trade-offs

- risk vs. safety
- cheap vs. expensive
- full functionality vs. reduced functionality

Dig Preserv class intro, 9/9/09

40

## A few questions our community should address

- Special issues raised by non-library institutions
- Special issues raised by images and rich media
- What is the work (or salient points we need to preserve)?
- Bring the arts communities (artist intent, BAVC) together with the preservation repository communities and the preservation metadata communities
- Specifically get Cult Heritage communities involved with the selected OCLC/RLG recommendations
- Get cult heritage groups started on working to make sure that structure standards incorporate our works
- What organizations will take responsibility to save today's digital "ephemeral" materials (online 'zines, arts discussion groups, etc.)?

Dig Preserv class intro, 9/9/09

41

## Standards, Metadata, & Best Practices to follow-

- Risk Management
- Best Practices for Reformatting
- Preservation Repositories & Metadata
- Other Metadata & Standards

Dig Preserv class intro, 9/9/09

42

## Risk Management

- We can't say definitively that we can make every digital work persist
- What we CAN say is that the more a digital work conforms to standards and best practices, the greater the likelihood that we can assure persistence
- Our preservation repositories can even accept deposits of non-conforming works, but the less they conform, the less likely that they'll be salvageable
- Persistence is most likely for works that share standards, metadata, and best practices

Dig Preserv class intro, 9/9/09

43

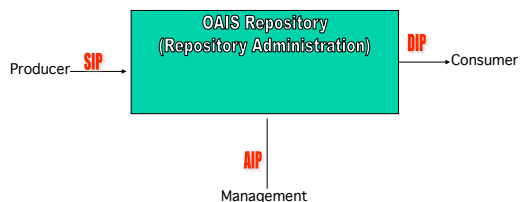
## Reformatting Best Practices (still images)

- Think about users (and potential users), uses, and type of material/collection
- Scan at the highest quality that does not exceed the likely potential users/uses /material
- Do not let today's delivery limitations influence your scanning file sizes; understand the difference between digital masters and derivative files used for delivery
- Many documents which appear to be bitonal actually are better represented with greyscale scans
- Include color bar and ruler in the scan
- Use objective measurements to determine scanner settings (do NOT attempt to make the image good on your particular monitor or use image processing to color correct)
- Don't use lossy compression
- Store in a common (standardized) file format
- Capture as much metadata as is reasonably possible (including metadata about the scanning process itself)

Dig Preserv class intro, 9/9/09

44

## Preservation Repositories: Open Archival Info System Model



Dig Preserv class intro, 9/9/09

45

## Preservation Repositories: Open Archival Info System Model

- ✦ High-level reference model describing submission, organization and management, and continuing access
- ✦ Conceptual framework for different organizations to share discussions with a common language
- ✦ Producers, consumers, management, actual repository
- ✦ SIP, DIP, AIP
- ✦ AIP consists of data objects plus representation info (Content, Preservation Description, Packaging, Descriptive)
- ✦ Originally developed for Space Science community

Dig Preserv class intro, 9/9/09

46

## OCLC/RLG Digital Repository Attributes

- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological suitability
- System security
- Procedural accountability
- Certification-

Dig Preserv class intro, 9/9/09

47

## Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist

<http://www.crl.edu/content.asp?11=13&12=58&13=162&14=91>

- Organizational Infrastructure
- Digital Object Management
- Technologies, Technical Infrastructure, & Security
- Who can benefit?
- The document benefits those who work in or are responsible for digital repositories and who want to be certified against its requirements, as well as for those who execute the audit and certification process. However, the guidelines are also designed for a much broader audience—for organizations planning repositories and for producers of digital materials.
- When does it come in handy?
  - For staff working within repositories: The checklist is useful in helping them to determine the criteria they need to plan for and to meet for safely storing and migrating digital materials.
  - For producers of digital materials: It also serves producers of digital materials by helping them know what to expect from the repositories they work with and how to streamline interactions with the repositories.
  - For those considering outsourcing their digital materials storage: The checklist will help organizations considering an external agency to serve their digital preservation needs to ensure that contracted organizations are executing the task adequately.

Dig Preserv class intro, 9/9/09

48



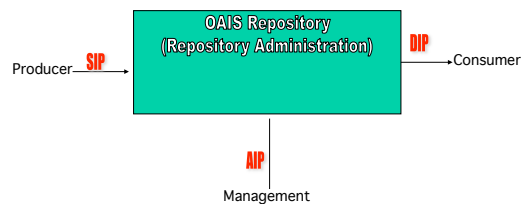
## OCLC/RLG Selected Recommendations

- Policies, Certification processes, Risk management, Persistent ID, Migration/Emulation experiments
- Stakeholders meet to decide how to describe what is in a dig repository
- Examine special properties of particular classes of digital objects
- Technical standards for exchange and interoperability btwn repositories
- Develop projects and case studies
- Copyright issues

Dig Preserv class intro, 9/9/09

49

## Management: Preservation Repositories: OAIS & agreements



Dig Preserv class intro, 9/9/09

50

## OAIS-steps in the process

- Agreement
- Ingest
- Store and manage/maintain
  - Refresh
  - Emulate/Migrate/others
- Disseminate

Dig Preserv class intro, 9/9/09

51

## Metadata

- Containers/Packaging for SIP (METS)
- AIP
- Preservation (PREMIS)

Dig Preserv class intro, 9/9/09

52

## AIP Metadata

- Preservation Description Info
  - reference info
  - context info
  - provenance info
  - fixity info
- Packaging Info
- Descriptive Info
- Content Info

Dig Preserv class intro, 9/9/09

53

## OCLC/RLG Efforts Working Group I: Preservation Metadata Framework

- ...to define the concept of preservation metadata, describe its importance in context of the overall digital preservation process, examine the "state-of-the-art" in the use of metadata in support of digital preservation, and evaluate the prospects for a community-wide, consensus-building activity in the area of preservation metadata (Preservation Metadata for Digital Objects: A Review of the State of the Art [http://www.oclc.org/research/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/pmwg/presmeta_wp.pdf))
- ...to develop a framework outlining the types of information—i.e., metadata—that should be associated with an archived digital object. (A Metadata Framework to Support the Preservation of Digital Objects [http://www.oclc.org/research/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/pmwg/pm_framework.pdf))
  - an expanded conceptual structure for the Open Archival Information System (OAIS) information model, and
  - a set of metadata elements, mapped to the conceptual structure and reflecting the information concepts and requirements articulated in the OAIS model.

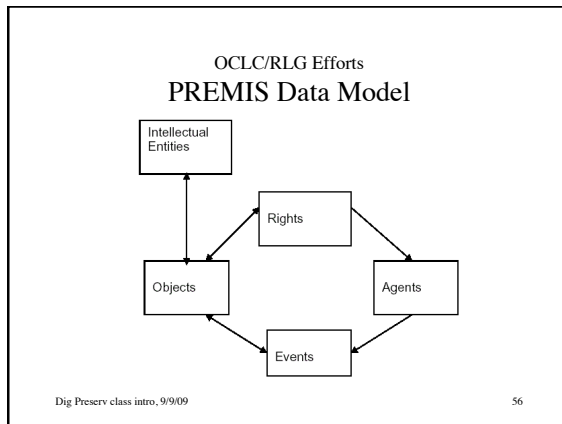
Dig Preserv class intro, 9/9/09

54

OCLC/RLG Efforts  
**Working Group II: PREservation  
 Implementation Strategies (PREMIS)**

- develop a core set of implementable preservation metadata elements, with broad applicability within the digital preservation community
- develop a data dictionary to support the preservation metadata element set
- examine and evaluate alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system, as well as for the exchange of preservation metadata between systems
- develop a pilot program for testing the group's recommendations and best practices in a variety of systems settings
- explore opportunities for the cooperative creation and sharing of preservation metadata

Dig Preserv class intro, 9/9/09 55



OCLC/RLG Efforts  
**PREMIS Data Dictionary Example**

|                                   |   |             |                             |
|-----------------------------------|---|-------------|-----------------------------|
| <b>Semantic unit</b>              | fixity  |             |                             |
| <b>Semantic components</b>        | messageDigestAlgorithm, messageDigest, messageDigestOriginator  |             |                             |
| <b>Definition</b>                 | Information used to verify whether an object has been altered in an undocumented or unauthorized way.   |             |                             |
| <b>Data constraint</b>            | Container   |             |                             |
| <b>Object category</b>            | <b>Representation</b>   | <b>File</b> | <b>Bitstream</b>            |
| <b>Applicability</b>              | Not applicable (see usage note)   | Applicable  | Applicable (see usage note) |
| <b>Repeatability</b>              |   | Repeatable  | Repeatable                  |
| <b>Obligation</b>                 |   | Optional    | Optional                    |
| <b>Creation/Maintenance notes</b> | Automatically calculated and recorded by repository.  |             |                             |
| <b>Usage notes</b>                | <p>To perform a fixity check, a message digest calculated at some earlier time is compared with a message digest calculated at a later time. If the digests are the same, the object was not altered in the interim. Recommended practice is to use two or more message digests calculated by different algorithms.</p> <p>The act of performing a fixity check and the date it occurred would be recorded as an Event. The result of the check would be recorded as the eventOutcome. Therefore, only the messageDigestAlgorithm and messageDigest need to be recorded as objectCharacteristics for future comparison.</p> |             |                             |

Dig Preserv cla... 57

- Digital Preservation Players**
- Collection staff (need to reach agreement on SIP/DIP and acceptable AIP transformations)
    - preservation/conservation staff
    - metadata staff
    - access staff
  - Repository staff
  - Agreement negotiators
- Dig Preserv class intro, 9/9/09 58

- Preservation Repositories:  
 too difficult for small institutions**
- Too complex for small institutions to manage
  - Will be done through partnering (small museum or dance company with University) or through consortia (museum association, state-wide organization, ...) or through service bureaus (OCLC)
  - Archive or museum will direct what is needed, but digital repository will carry out the actual work (as defined in SIP/DIP/AIP agreement)
- Dig Preserv class intro, 9/9/09 59

- Many quality questions**
- Quality of playback?
  - Theater experience?
- Dig Preserv class intro, 9/9/09 60

## Structural & Administrative Metadata (METS)

- For complex, multi-part works (collections of images, groups of maps, albums of photographs, etc.)
- Structural--preserving relationships (books/chapters /sections/pages, photo albums on a page, master images with thumbnails and different sizes, etc.)
- Administrative--keeping information about managing the works over time (keeping track of all of the pieces in a group, what software is needed to decompress/view an image, etc.)
- Metadata Encoding & Transmission Standards (METS)-  
-wrap the image(s) in XML encoding

Dig Preserv class intro, 9/9/09

61

## Paradigms Shifts needed

|                              | <u>Old</u>               | <u>New</u>                                    |
|------------------------------|--------------------------|---|
| <b>Physical preservation</b> | atmospheric cntrl        | ongoing mgmt                                  |
| <b>What to save?</b>         | artifact                 | idea + ancillary material & documentation     |
| <b>Cataloging</b>            | Individual work in hand  | FRBR  |
| <b>Later access</b>          | Artifact & documentation | Restaging, ancillary material & documentation |

Dig Preserv class intro, 9/9/09

62

## Introduction to Digital Preservation: Issues, Principles, Players, & OAIS

Howard Besser, NYU Moving Image Archiving & Preservation Program

- <http://besser.tsoa.nyu.edu/howard/longevity/>
- <http://www.nyu.edu/tisch/preservation>
- <http://www.thirteen.org/ptvdigitalarchive/>
- <http://www.oclc.org/research/projects/pmwg/>
- <http://www.interpares.org>
- <http://www.digitalpreservation.gov/>
- UC Libraries Systemwide Operations and Planning Advisory Group (SOPAG) Site  
<http://www.slp.ucop.edu/sopag/> for the UC Digital Preservation & Archiving Committee Final Report
- <http://variablemedia.net/>
- [http://www.firstmonday.dk/issues/issue7\\_6/besser/](http://www.firstmonday.dk/issues/issue7_6/besser/)
- <http://sunsite.berkeley.edu/Metadata/sp2000.html>
- <http://www.niso.org/commitau.html>
- <http://www.ifla.org/II/metadata.htm>
- METS official site: <http://www.loc.gov/standards/mets>