In a Digital World, Image & Sound Collections require Stewardship: Selection, Curation, Preservation, Presentation

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

Digital Stewardship definitions
• “acquire, manage, organize, preserve, and provide access to massive amounts of data for use and re-use by a variety of interdisciplinary and heterogeneous communities over time”
• “Preservation + Use”

Digital Stewardship:
• Includes much of what cultural institutions have always done (select, acquire, preserve, provide access, maintain context)
• But also adds some of the activities that audiovisual collections have always done (maintain knowledge of the original equipment and formats, manage process of reformatting when older hardware is difficult to maintain)
• Also includes important notions from the digital world about ongoing management of re-formatting information, technical metadata, contextual info, repository issues, …

What do Image & Sound Museums already do?-
• Have broad collections where each needs to be handled in different ways
• Handle Re-Formatting
• Have an online presence & provide access to their collections
• Encourage new activities in the Sound & Image communities
And other cultural organizations have similar collections (but sometimes for different purposes)
Handle Re-Formatting

MIS Collection Content: video transfer in RJ

Have an online presence & provide access to their collections

MIS online activities-RJ

MIS online activities-SP

MIS online activities: Highlighting Collections
Encourage new activities in the Sound & Image communities

MIS activities in Campinas

What kind of new things do we need to learn about?
- Digital Preservation
- Digital Curation
- Digital Selection

The Short Life of Digital Info: Digital Longevity Problems-
- Digital Preservation
  - The Viewing Problem
  - The Scrambling Problem
- Digital Curation
  - The Inter-relation Problem
  - The Custodial Problem
  - The Translation Problem

Digital Preservation
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

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

The Scrambling Problem

Dangers from:
- Compression to ease storage & delivery
- Encryption to enhance digital commerce

Save the Hardware & Software-

- A huge undertaking
- Computer Museum
- Broderbund

Keep Old Video Equipment--NARA

Keep Old Video Equipment--NARA
Keep Old Audio Equipment-NARA

Old Video Formats

Old Digital Formats

Save the Hardware & Software

• A huge undertaking
• Computer Museum
• Broderbund

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

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
**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?)

---

**Brittle Newspapers** (Australia Battye Library)

---

**Film Decay**

---

**Why do we Reformat?**

- Because we cannot sustain the original object (its physical characteristics are deteriorating too fast)
- Because continued access and handling of the original object will rapidly decay its physical characteristics (so we create a surrogate for users and store the original in very good conditions, away from users)
- Because viewing the work requires some kind of technology, and we can’t keep that technology working very far into the future-

---

**Lost Tapes, Found Sounds Exhibition**

Harold Schellinx

---

**Obsolete or deteriorated Physical Carriers**
Limitations of Reformatting

- Authenticity issues
- User behaviors (newspaper, book, video game, …)
- Users mistaking the reformatted work for the original
Critiques of Reformatting

Mainly User Behaviors

• Can’t view outside the library
• Only sequential access
• Viewing and studying is awkward
• …

But if we don’t Reformat, we totally lose some kinds of works
(particularly audiovisual works like film)

• 50% of all titles produced before 1950 have vanished (approximate number as of late 1970s)
• This reflects full-length features; survival rates are much lower for other types (studio newreels, shorts, docs, independent, …), and these “orphans” are particularly in peril
• Fewer than 20% of features from 1920s survive in complete form; survival rates of 1910s is <10% (& none of these are negatives)


And sometimes we have to reformat because of technology changes

• We don’t have video players to play tapes made 25 years ago
• We don’t have 8-inch floppy disk drives, syquest drives, zip drives
• We don’t have Windows 3 operating systems
• But this is something that conservators have always dealt with…

We sometimes have no control over the technologies we use

• Environmental--Government mandates to discontinue use of halon gas as fire extinguisher
• Economic--Companies in economic trouble will cease manufacturing technologies that aren’t hugely profitable
  – Ilford (2004 bankruptcy)
  – Agfa (2005 bankruptcy)
  – Kodak

Kodak stops making some films

Kodak stops making some papers

Kodak stops making some films
Basic Economics

- The cultural community is not large enough a purchaser to sustain many types of manufacturing and technological production
- Many of the things we use are based upon larger production runs for larger (and richer) communities
- Therefore, we need to be periodically monitoring the economic health of our suppliers, and be aware of long-term trends affecting their other customers

Storing on CDs becomes a big problem over time

Consumers replace their CDs with a hard disk (& so should you)

Possible endless need for reformatting implies

- Possible loss with each generation
- Requires managed environment

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

Paradigms Shifts needed

<table>
<thead>
<tr>
<th></th>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical preservation</strong></td>
<td>atmospheric cntrl</td>
<td>ongoing mgmt</td>
</tr>
<tr>
<td><strong>What to save?</strong></td>
<td>artifact</td>
<td>idea + ancillary material &amp; documentation</td>
</tr>
<tr>
<td><strong>Cataloging</strong></td>
<td>Individual work in hand</td>
<td>FRBR</td>
</tr>
<tr>
<td><strong>Later access</strong></td>
<td>Artifact &amp; documentation</td>
<td>Restaging, ancillary material &amp; documentation</td>
</tr>
</tbody>
</table>
Digital Curation

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)?

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?

The Custodial Problem: How to save information?
- Methods for later access
  - Refreshing
  - Migration
  - Emulation
- Issues of authenticity and evidence

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

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
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-

Screen Formats

Many quality questions

- Quality of playback?
- Theater experience?

Standards, Metadata, & Best Practices to follow-

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

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

Reformatting Best Practices

- Think about users (and potential users), uses, and type of material/reflection
- Scan at the highest quality that does not exceed the likely potential uses/uses intended
- 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 grayscale 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)
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

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

- Organizational Infrastructure
- Digital Object Management
- Technologies, Technical Infrastructure, & Security

OAI Functional Entities

- OAIS-steps in the process
  - Agreement
  - Ingest
  - Store and manage/maintain
    - Refresh
    - Emulate/Migrate/others
  - Disseminate

OCLC/RLG

- Digital Repository Attributes
  - Administrative responsibility
  - Organizational viability
  - Financial sustainability
  - Technological suitability
  - System security
  - Procedural accountability
  - Certification
Metadata

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

AIP Metadata

• Preservation Description Info
  – reference info
  – context info
  – provenance info
  – fixity info
• Packaging Info
• Descriptive Info
• Content Info

OCLC/RLG Efforts

PREMIS Data Model

Intellectual Entities

Rights

Objects

Agents

Events

OCLC/RLG Efforts

PREMIS Data Dictionary Example

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data model</td>
<td>PREMIS</td>
</tr>
<tr>
<td>Object type</td>
<td>Event</td>
</tr>
<tr>
<td>Representation</td>
<td>File</td>
</tr>
<tr>
<td>Provenance</td>
<td>Unknown</td>
</tr>
<tr>
<td>Authenticity</td>
<td>Unknown</td>
</tr>
<tr>
<td>Reliability</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Digital Curation Life-cycle

http://www.dcc.ac.uk/

Digital Curation

• Planning & risk management
• Open flexible file formats
• Format registries
• OAIS repositories
• Extensive Metadata
• Individuals & organizations-
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

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)

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?

Selecting and managing digital works

- “born-digital” works
  - Web-based media (like the first songs of a band that later becomes famous)
  - Email, Documents, Architectural drawings, Photographs
  - Inventory records & DBs
  - CD ROMs
  - DVDs
- Works converted to digital form
  - Digitized documents
  - Digitized photos
  - Digitized maps
  - Digitized audio
  - Digitized video
  - …
We convert works to digital or need born-digital because:

• We want to make documents or photographs available to people without them having to come to our building
• We want to capture documentation of our intangible heritage (patrimonio imaterial) (music, dance, performance)
  – And most of today’s documentation methods are digital (sound, photographs, movies)

SP Supreme Court Justice Enio Santarelli Zuliani shuts down YouTube

Out of Print

Other interesting projects

• Digital Public Television Preservation
• Disaster Prevention & Recovery
  – Restoration and distribution of Helen Hill’s works
  • …
NYU/Public Television Project

- Preserve a broad set of elements (including ancillary material)
- Life-cycle mgmt (add metadata as soon as a clip comes in)
- Establish a community of stakeholders, working together for preservation (stations, university, librarians, journalists, historians, producers, scholars, …)
- Build an OAIS Server
- Explore appropriate file formats, wrappers, METS extensions
- Develop sustainable business model

Public Television is home of

Local broadcasting is nearly all via digital systems

Thirteen Master Control manages one high definition, two analog and three digital over-the-air broadcast channels.

The key is designing a preservation repository that the system can afford to maintain and use.

To test this out, we are focusing on --

- Appraisal and Selection – developing criteria and standards for what to preserve and by whom
- File Formats and Packages – determining the best formats for our various uses, plus testing the suitability of file “packaging” for long term preservation
- Metadata and Related Topics – specifying technical, descriptive and rights information
- Repository Design – technical architecture, administrative policies and potential business models
- Sharing Our Findings – Keeping the public broadcasting community involved and informed all along the way

Project activities include --

- Completing an inventory of at-risk materials to better quantify our holdings and prepare for selection
- Reviewing best practices and most up-to-date developments in the field of video archiving
- Conducting facilitated discussions on key topics to guide setting standards and policies
- Establishing an Advisory Committee to assist with selection criteria
- Ingesting sample materials and testing the repository
- Presenting regular reports to public broadcasting and moving image archive community for ongoing feedback

http://www.nyu.edu/tisch/preservation/research/
New Orleans: Film Projector

New Orleans: Still Photo Damage

New Orleans Films projected
Helen Hill films

New Orleans Films projected
Helen Hill films

New Orleans Films projected
Helen Hill films

New Orleans Films projected
Helen Hill films
We looked at preventative techniques & their limitations

Conclusions for preserving all types of digital works:
Digital Repository Traditions & Services require

- Sustainability
- Interoperability
- Access
- And all of these require Standards and Metadata

Conclusions for preserving all types of complex works:
From the technological point of view
Standards offer the best hope of overcoming Impediments

- Easier to maintain a single set of standards over long periods of time
- Puts your institution in the same large boat with lots of other institutions who will face obsolescence and migration problems periodically throughout the future

for artistic and other challenging works:
How Best to save these works?

- Use Standards wherever possible
- Be aggressive about asset mgmt -- saving component parts and ancillary materials
- Both creator and Archive should develop an institution-wide plan for saving electronic works
  - Refreshing and either migration or emulation
  - Standard encoding schemes
  - What is the work? And prioritize what needs to be saved
  - Save ancillary materials and records
Paradigms Shifts needed

<table>
<thead>
<tr>
<th>Physical preservation</th>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>atmospheric ctrl</td>
<td>ongoing mgmt</td>
</tr>
<tr>
<td>What to save?</td>
<td>artifact</td>
<td>idea + ancillary material &amp; documentation</td>
</tr>
<tr>
<td>Cataloging</td>
<td>Individual work in hand</td>
<td>FRBR</td>
</tr>
<tr>
<td>Later access</td>
<td>Artifact &amp; documentation</td>
<td>Restaging, ancillary material &amp; documentation</td>
</tr>
</tbody>
</table>

So, with electronic works, the focus should be less on stable temperature (Helsinki underground vaults)

And less on the construction of Vaults (Helsinki underground vaults)

But more on ongoing management of a work without worrying so much about physical embodiment

In a Digital World, Image & Sound Collections require Stewardship: Selection, Curation, Preservation, Presentation

Howard Besser, NYU Moving Image Archiving & Preservation Program

- http://besser.tsoa.nyu.edu/howard/Talks/
- http://besser.tsoa.nyu.edu/howard/longevity/
- http://www.nyu.edu/tisch/preservation
- http://www.nedcc.org/resources/digitalhandbook/ix.htm
- http://www.dcc.ac.uk/
- http://www.digitalpreservation.gov/
- http://www.ifla.org/II/metadata.htm
- METS official site: http://www.loc.gov/standards/mets