Managing the Intangible

By Steven Puglia

Preservation and Imaging Specialist
U.S. National Archives and Records Administration
8601 Adelphi Road, Room B810
College Park, MD 20740, USA
Phone: 301-837-3616
Email: steven.puglia@nara.gov

April 2007

We no longer have a choice - compared to 4 years ago:

- Conventional methods of analog reformatting are becoming unavailable or obsolete for many types of records and formats.

- User expectations have changed as well - people expect all historic resources to be online.

- We are and will have to use digitization for preservation reformatting of all types of records.

We can not ignore digital technology, it is not going to go away.

We must learn to:

- Be critical observers of technology.

- Honestly evaluate its strengths and weaknesses.

- Be sophisticated users of all appropriate tools.

Static Analog Media

Printed publications, photographs, microfilm, and digital hardcopy output.

Dynamic Analog Media

Motion pictures, audio recordings, and video recordings.

Digital Media

All types.

Static Analog Media

Chemical and physical stability limits usable life.

Dynamic Analog Media

Chemical and physical stability and systems obsolescence limit usable life.

Digital Media

Systems obsolescence is critical factor, but chemical and physical stability can be problematic.

Digital Media

Digital

*Dynamic
Analog -
machine
dependent*

Motion Pictures

Video Recordings

Audio Recordings

Increasing
complexity
and

corresponding
risk of loss.

*Static Analog -
human readable*

Still Photographs

Textual Records

Digital - digital preservation

Motion Pictures

Video - digital ref. only option

Audio - digital ref. only option

Harder to do preservation reformatting as digital - but once digital, potentially no loss when copying.

Still Photographs

Textual Records

Feasible to do preservation reformatting as digital.

Increasing expense to reformat and to preserve.

For preservation - use all of the following tools:

- Cost-benefit analyses

- Proper Storage / Cool and cold storage

- Environmental monitoring

- Holdings maintenance and conservation

- Risk, condition, and archival / curatorial assessment

- Reformatting

- Reformatting is almost always the most expensive option.

- For records that benefit from cold storage and are not susceptible to format obsolescence, we can store originals in cold storage for between 100 to 200 years for the same cost as reformatting everything.

- Reformatting remains the primary approach for preserving many system dependent formats - particularly audio and video recordings.

- Reformatting remains essential for facilitating access to all formats.

- Reformatting allows originals to remain in secure and/or cold storage.

Institutional and User Perspective

From Steve Chapman's "Microfilm: A Preservation Technology for the 21st Century?"

- Use

- Sustainability

- Affordability

Original Resource Perspective

From Steve Puglia's Reformatting pyramid chart

- Fidelity

- Functionality

- Longevity

- Source characteristics

- Purpose/People

- Technology capabilities

From Paul Conway's "Overview: Rationale for Digitization and Preservation"
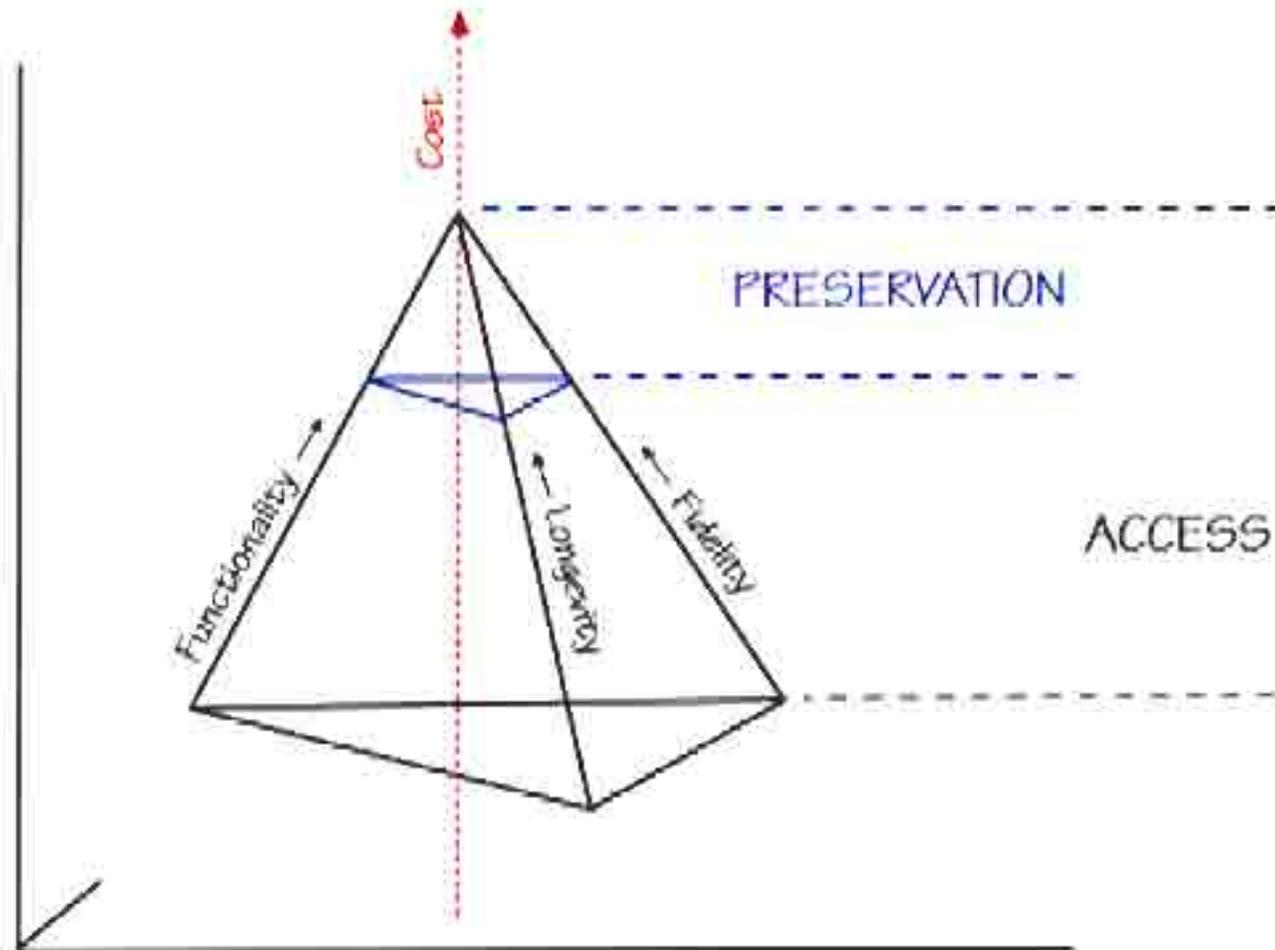
REFORMATTING OPTIONS

Diagram by Steven Puglia, Oct. 1998

Digital reformatting is not necessarily easier or less expensive than traditional approaches to analog reformatting.

- The amount of overall work is comparable, digital is often slower, and the technical complexity in many cases is greater.

- For many types of originals, reformatting labs will need multiple digital workstations to achieve the same capacity as traditional analog lab equipment.

Digitization equipment and supporting information technology (IT) infrastructure will have to be replaced much more frequently.

- Analog lab equipment often had a usable life of 10 to 30 years, and sometimes longer.

- Digital equipment will need to be replaced every 3 to 5 years.

Analog is not going away any time soon-

- •We will have to maintain analog infrastructure - particularly for machine-dependent formats - audio, video, and motion pictures.

- •Not so easy as analog equipment is discontinued and ages - maintenance, parts, and repair will not be easy.

We will continue with analog reformatting in certain areas as long as possible-

•Like motion pictures - because there are technical and economic limitations with digital approaches.

•For motion pictures it is easy to convert to digital video at lower resolutions, but hard to do preservation reformatting.

We will have to incur major reformatting expenses to copy machine-dependent formats in a timely manner.

It is not always possible or practical to speed up the copying process when the originals are in analog format.

Need to take advantage of the benefits of computers and information technology.

Need to develop comprehensive workflow tools and automate as much as possible to gain efficiencies in digitization on a large scale.

Digitization is equal parts-

- Digital conversion

- Description / cataloging / indexing and creation of other metadata

- Project management

Systems Perspective:

Managing and preserving digital data/objects/records is different than managing and preserving physical records.

Digital

- Everyone working to determine all that will be necessary to preserve digital data over the long term - we have made tremendous progress over the last 5 to 10 years.

- Will need to preserve digitally created materials using digital technology.

- Media reversion is not an option for many types of records, will lose functionality and then it is no longer the record.

Need to move away from managing and preserving technology to approaches for managing and preserving digital objects/data.

Actively manage risk.

Need to bring digital resources into a managed environment as soon as possible, to facilitate management, access, and long-term preservation.

In order to consider using digitization as a method of preservation reformatting -

It will be necessary to specify more about the characteristics and quality of the digital copies than what is in many current digitization guidelines.

As an example - for raster images, need for consensus on a variety of technical factors such as:

- Capture device performance
- Tone reproduction and orientation
- Color reproduction
- Bit depth
- Sampling frequency or spatial resolution
- Other quality parameters
- Image and signal processing procedures

•NARA's 2004 *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images* may prove appropriate for certain collections and institutions.

•But - the decision to use the *Guidelines* should be based on a critical evaluation of needs, and not just be accepted because the *Guidelines* are available.

Other aspects beyond imaging parameters must be addressed in order to be considered preservation reformatting.

- •Such as selection, metadata, rights, digital preservation, managed environment, etc.
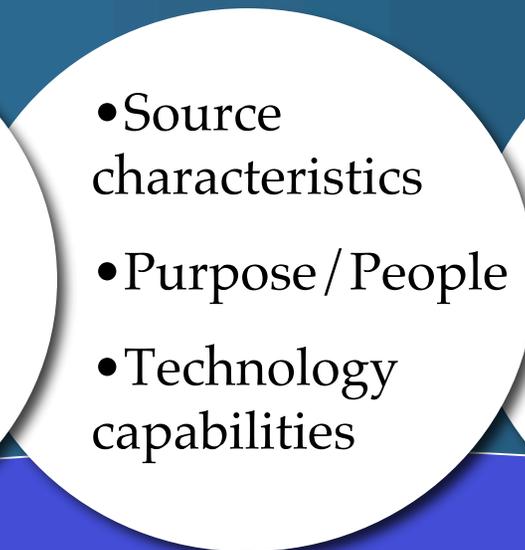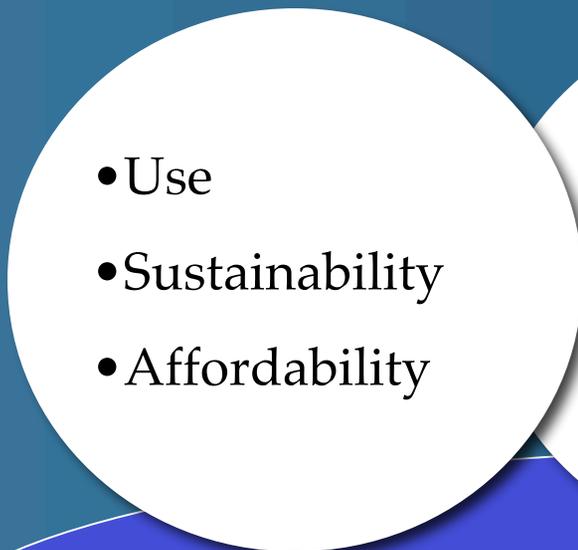- •Currently, the NARA *Guidelines* do not address these issues.

Why use digitization for preservation reformatting?

- Appropriate analog processes and materials are not or will not be available in the future for reformatting originals.

- We will have to use digitization for all preservation reformatting, just as audio recordings now can only be reformatted digitally.

- Digitization meets current needs for facilitating access and limiting handling of originals.

Defining approaches for preservation reformatting using digitization -

Started work two years ago on defining levels of information capture for photographs.

Imaging Environment

Defined ⟷ Undefined

Image State

Prepped for a specific output ⟷ RAW

⟷

Output Referred - looks towards output

Input Referred - looks towards sensor

⟷

Original Referred - defined relationship between original and digital version

⟷

Current Practice

Emerging Practice

Should be able to get by with less technical metadata ⟷ More technical metadata is needed

Follow appropriate technical guidelines and specifications, based on explicit decisions relating to specific needs.

Need to collect all appropriate metadata - documenting originals and digital objects.

Digitizing guidelines exist - still work to be done:

- Textual records - original paper records and microfilm
- Still photographs
- Audio recordings
- People working on video recordings and motion pictures

Need to define essential characteristics of the original resources -

- •Microfilm standards/guidelines focus on maintaining text legibility.

- •Specifications for photographic duplicates define approaches to produce duplicates that have the same photographic properties as the originals- same overall density, density range, and relationship between the tones.

Other Issues-

- Quality control (QC) and quality assurance (QA).
- Scanner/digital camera and digitization equipment evaluation and performance - limits and variability.

High-quality A/D converter (circa 2006):

- Frequency response (1 Hz to 48 kHz): -1 dB
- Total harmonic distortion and noise (1 kHz at 0 dBFS): < -108 dB (0.0005%) (unweighted RMS)
- Dynamic range: > 130 dB (unweighted RMS)
- Intermodulation distortion: <-90 dB
- Spurious aharmonics: <-130 dBFS
- Crosstalk (50 Hz, 0 dBFS in opposite channel): < -130 dB
- Crosstalk (15 kHz, 0 dBFS in opposite channel): <-140 dB
- Linearity (at -144 dBFS): < 3 dB
- Intrinsic jitter: < 18 pSec RMS
- Phase linearity: < 1。
- Internal clock accuracy: ±5 ppm

"Measurement and Evaluation of Analog-to-Digital Converters Used in the Long-term Preservation of Audio Recordings" by Ken Pohlmann, www.clir.org/activities/details/AD-Converters-Pohlmann.pdf

For traditional reformatting, we developed infrastructure and defined workflows.

Same for digital reformatting.

For digital reformatting, need at least three tiers of infrastructure-

- <u>Highest</u> - Digital Repository

- <u>Intermediate</u> - Shared Work Environment

- <u>Lowest</u> - Digital Reformatting Lab

At all levels - need to do the routine IT things to mitigate risk of data loss.

It is about risk management.

It is about creating a managed environment – the need to be proactive, not just reactive.

Effective IT procedures exist for the short-term management of electronic records and digital information-

- Not always followed.

- Not always as easy or as inexpensive as advertised.

- We have been sold on the promise of the technology, but rarely acknowledge the downsides.

Digital Reformatting Laboratory

The local lab infrastructure -

- Not about managing and making information available.
- It is about creating digital objects from original records of all types.
- The digital equivalent of the traditional reformatting lab.
- Perform digitization, quality control, collect technical metadata, and in some cases perform baseline indexing.

At the lab level - need to define:

- Baseline workflows to enhance productivity.
- More specialized or custom workflows to address problems with original resources and to accommodate exhibits and publications.

Shared Work Environment

- Short-term place to make resources available to complete all other work.
- Place where all additional work can be done after the digital objects are created - such as additional QC, metadata creation and indexing, QC on metadata, etc.

Digital Repository

- Long term management and preservation of digital copies.
- For NARA this will be the Electronic Records Archives (ERA).

The infrastructure will assist with-

Project / Reformatting Prep Work

1. Manage workflow and coordinate preservation activities among archival units, preservation labs, exhibits and publications staff, web staff, and access staff.

2. Prepare documents for digitization, both archival and preservation/conservation preparation.

3. Collect and record a more detailed level of descriptive metadata during the course of archival prep work and to enhance the authoritative archival descriptions in official descriptive metadata repository.

4. Document preservation/conservation actions in preparation for digitization.

Reformatting Lab Work

1. Perform digitization or analog (photographic duplicate) reformatting.

2. Perform digitization using local workstations with specialized equipment and software, and with local lab network including local digital storage capacity to facilitate lab activities.

3. Collect appropriate non-descriptive technical metadata resulting from the digitization process, like specifications followed to produce digital resource.

4. Perform preliminary organization of the digital resources.

Post-Lab or Digitization Work

1. Organize/structure digital resources based on the intellectual and physical organization of the records.

2. Automated quality control (system checks) in addition to staff quality control (human checks) to identify rework/defect correction.

3. Rework/defect correction efforts.

4. Track and associate new digital/analog versions of records with the original records.

5. Collect and prepare digital resources and metadata about the resources for export into other systems.

6. Provide for long-term digital storage for digital assets and related metadata.

Creating-

- Lots of choices available for digitizing equipment - even inexpensive office scanners can do a good job.

- Scaling to meet large production demands, metadata creation/collection is a problem, and keeping up quality (both resources and metadata) is a problem.

- There are efficiencies when working from existing intermediates like microfilm, but may introduce deficiencies and may not capture all the desired characteristics of the originals.

Creating-

- While it seems easy to do, digitization still requires a lot of technical knowledge and skill

- Can not ignore all the other things that need to be done - like metadata creation - good descriptive information is even more critical for the accessibility of and long-term preservation of digital copies.

Storage-

- Moving away from individual removable storage media (discs or tapes) to storage on spinning discs (servers with RAID).
- Do not want to be managing physical media - scalability of approach and management is a problem.

Retrieving-

- Metadata - standards and what is the minimum?

- IT infrastructure - hardware and software.

- Software - applications sold to many market segments that can be used.

Our need for software systems does not match a specific model or market segment, includes the following functionality-

- Records management application (RMA)
- Content management (CM)
- Document capture and management
- Digital asset management (DAM)
- Media asset management (MAM)
- Digital repository
- Digital preservation

The cost for digital storage continues to drop, but the managed environment (digital repository) may not be so affordable - compared to storing and managing human-readable and analog copies

Options-

- Do everything locally - build and manage the infrastructure yourself

vs.

- Cooperative/consortial approaches

vs.

- Shared commercial options - hosted services and application service providers

Conclusions:

- Start with getting the basics right, including hardware / software infrastructure and metadata requirements.

- Finish defining appropriate approaches for preservation reformatting using digitization.

- Use all appropriate tools - including digitization and new approaches for digital preservation.