

Microfilm: A Preservation Technology for the 21st Century?

*Stephen Chapman
Harvard University
Cambridge, Massachusetts, USA*

Abstract

The National Endowment for the Humanities, libraries, and library consortia have made significant investments in preservation microfilm during the past 20 years. Is it a viable preservation solution for the next 20? This is an opportune time for technical experts to revisit functional requirements for preservation and to state them in a format-neutral fashion, for use in informing the development of preservation standards, guidelines, products, procedures, and services.

Criteria for *use*, *sustainability*, and *affordability* are presented as a draft framework to engage IS&T and other interested professional associations in a discussion of variables relevant to codifying preservation reformatting processes, and to certifying copies according to preservation standards.

Paper at Risk, Perspectives: 1985-2005

In 1985, the National Archives and Records Administration (NARA) estimated that 530 million of their 3 billion paper documents were at high risk of loss. Action was needed, but *which* action(s) would be most effective? NARA solicited recommendations from the National Materials Advisory Board of the National Research Council, which convened a distinguished technical committee to assess a range of factors contributing to paper deterioration. Charged to investigate methods to transfer “information from original paper records to media having acceptable permanence,”¹ the Committee on Preservation of Historical Records may have been the first to evaluate paper, microfilm, and magnetic tape as candidate formats for “archival copying.” They concluded that standard electrophotographic processes and silver-based micrographic processes were the two methods “appropriate for archival preservation.”

In 1988, a broad coalition of stakeholders in preserving U.S. cultural heritage testified before Congress in the special hearing on the appropriation for the National Endowment for the Humanities (NEH) to raise awareness of the “brittle book crisis.” George Farr, Jr. and Deanna Marcum’s engaging history of the evolution of preservation programs from the early 1980s to 2001 recounts how the Commission on Preservation and Access, through the “Hayes Study” (1986), gathered and presented evidence that as many as 11 million uniquely held volumes in the nations’ libraries were at risk,

and that it was economically feasible to microfilm at least 3 million of these volumes. At the time, no federal funds were being allocated for preservation. When asked how NEH would use an “enhanced Federal allocation” to address the problem, their Office of Preservation presented a multi-faceted plan to coordinate a distributed effort to preserve the intellectual record. Several of the assumptions underlying this plan became core principles for preservation reformatting:

- that preservation be undertaken in a manner to ensure broad access, including intellectual access (cataloging), to cultural resources whenever possible
- through research and demonstration, and education and training, that federal funds be used to support a distributed infrastructure of preservation programs, in addition to preserving endangered content
- that cooperative efforts should form “comprehensive collections” on topics extending beyond U.S. history and culture²

The 100th Congress voted to allocate \$153 million to NEH in the FY1989 budget, a \$12.5 million increase over the previous year. NEH’s plan defined a 20-year microfilming initiative, if allocations were maintained, that would preserve the intellectual content of up to 3 million volumes. To date, the NEH Division of Preservation and Access has awarded \$84,162,352 to 153 libraries and library consortia to preserve on microfilm 1,109,050 volumes.³

By 2004, some practitioners long dedicated to the enterprise of reformatting brittle books, and equally committed to exploiting the capabilities of digitization to expand access to collections, were strongly advocating that federal allocations for preservation support digitization as well as microfilming for preservation reformatting. Preservation librarians at the University of Chicago and the University of Michigan authored a report that the Association of Research Libraries (ARL) issued in June, 2004 as an endorsement of digitization as “an accepted preservation reformatting option.”⁴ This endorsement begs the question, “Should the remaining 1.9 million volumes targeted for preservation by the NEH Brittle Books Microfilming Program be microfilmed at all?” Given the benefits and viability of digitization, under which conditions does microfilm remain a viable or even preferred option?

Several weeks ago (December 14, 2004), Google's widely circulated press release announcing partnerships with major research libraries to digitize books—not just acidic books at risk, but in two noteworthy cases, *all* books—severely shook up the conventional wisdom regarding reformatting. Federal funders, libraries, library consortia, and scholarly societies have been responding to a crisis of brittle paper; Google sees the crisis of paper. Google's mission is “to organize the world's information and make it universally accessible and useful.” Where the ARL endorsement in 2004 challenged preservation librarians to reassess and reiterate the value of microfilm in a digital world, the Google announcement challenges librarians to consider the fundamental question of access to the intellectual record, “How can information be organized and made universally accessible if it is not in electronic form?”

What strategies do the attendees to the IS&T Archiving Conference envision being used during the next 20 years to preserve and broaden access to the intellectual content of books and other printed materials? What investments will their organizations make in technology and why? When confronted with challenges, solutions developers (as opposed to those who compare and contrast current solutions) first determine functional requirements.

What Does It Mean ‘To Preserve’?

Ask someone who has access to a networked computer what it means to make information widely accessible, and he or she is likely to respond, “Put it on the web.” For everyone else? Perhaps access is more nuanced and multi-faceted: *any* copy will do. Ask what it means to preserve the cultural record, what would the experts say? What principles and practices have been codified?

The *RLG Preservation Microfilming Handbook* (1992) has long served as the *de facto* set of best practices for preservation reformatting of printed material. The *Handbook* presents the case that conformance to good practice extends beyond capture (the act of producing a microfilm image), to film processing, packaging, storage, duplication, intellectual control, quality control at each stage, and, finally, distribution. Preservation is expressed by processes as well as products.

The RLG guidelines, however, are presented with an interesting caveat. The editor cautions that, while grounded in standards, the *Handbook's* recommended methods for microfilming, film processing, storage, and description represent only “best thinking” among experts, not assurances of long-term sustainability. The *Handbook* states, “Most of the guidelines are still subject to debate and fine-tuning, and will continue to evolve in the future... in the expanding national preservation effort and the evolutionary nature of the tools that support it.”⁵

The extensive use of digital imaging by libraries and archives in the past decade to improve access to printed materials has rekindled the debate and fine-tuning of guidelines, methodologies, and procedures. In part through this Archiving Conference, IS&T has positioned itself as an

arbiter of ideas, standards, and practices for document preservation.

This Archiving Conference presents a useful forum for technical experts to begin a discussion that might ultimately yield a set of *format-independent* criteria that can be used to evaluate preservation and access solutions according to the requirements that preservation must serve. Technologies evolve quickly; new formats emerge and make others obsolete. Does innovation, however, necessarily shift the preservation mandate?

As a means to promote discussion among the diverse international constituency of stakeholders in preservation drawn to this conference, the following outline presents a framework of criteria for *use*, *sustainability*, and *affordability* that, in concert, may be used to evaluate and adopt “best” strategies (the technologies, standards, and practices) for preservation reformatting of printed material.

Use Factors

A fundamental mission of libraries is to make information available for use. Use factors must take priority over sustainability and affordability, because regardless of how durable a copy may be—one thinks of the advertised 1,000 year lifespan of the High Density Rosetta (HD-Rosetta) from Norsam Technologies—or how inexpensive it may be to produce and maintain, a copy must be usable to be a *preservation* copy. Artifact preservation might serve as a fundamental underpinning to “preservation and access,” but handing a library patron an obsolete, uninterpretable object does not enable him or her to perform a given task. (In this library-oriented context, it is worth underscoring that human activity is analog: electronic data must be rendered in a human understandable fashion in order to be useful.)

Organizations that are highly service-oriented will take care to measure and respond to the needs of their core constituents. In this age of brick-and-mortar and digital libraries, each institution must be cognizant of whether its users will bother to use (“access”) material if it is not in the format they prefer.

Metrics in this category should enable practitioners to conform to minimal thresholds of best practice enabling people to locate and use copies of the published and archival works originally produced as printed material (e.g., books, pamphlets, serials, documents, manuscripts).

Object Identification

Preservation copies must be located in order to be used. What are the minimum criteria to sanction a copy as a ‘preservation copy’? Should the NEH-articulated mandate to ensure broad access be reiterated?

Specific arenas to mandate (or excuse) compliance include:

- Bibliographic description: Must preservation copies be identified in library catalogs? According to standards and best practices regarding data formatting, levels of description and access (e.g., subject-level description to

identify material by topic), authority control, and with notations of various holdings and versions?

- Registry compliance: Must preservation masters of published material—likely to be held by many libraries—be registered in databases to record preservation commitments?

Ensuring Persistence of Original Uses

When copies are made, particularly in formats that do not replicate the media and support inherent to the original paper and bindings of printed material, some intrinsic value will be lost. (Hence the commitment to Congress in the discussion of NEH appropriations to preserve the intellectual content of brittle books.) The overarching question to address in defining thresholds of compliance in preservation reformatting is, “Which uses inherent to the original material must persist in the copies delivered to the user?” By phrasing the question in this way, one can implement strategies in which preservation *masters* are not formatted to serve a single set of uses (e.g., to generate a near-facsimile printed codex), but formatted to be amenable to processing such that key attributes and uses of the original print may be replicated in one of the versions delivered to users.

For this category, compliance should apply at two levels for multi-page reproductions. For an item to be designated as a preservation copy, which thresholds must be met for:

- Quality of the Multi-Page Object: must this be expressed and measured according to standards and metrics for completeness, sequence (original order), and levels of navigation (enabling users easily to locate and turn to sections within the work)?
- Quality of Image Reproductions of Individual Pages: must this be expressed and measured according to absolute metrics for image quality (e.g., spatial resolution, noise, tone reproduction, and color reproduction); for legibility of meaningful components; for replication of dimensions, layout, age and condition?

Transformative Uses

To comply minimally with guidelines for preservation reformatting, must copies (preservation masters and/or delivery copies) facilitate transformative uses? In the modern world, must preservation masters support, or be capable of supporting uses not offered by printed originals?

The obvious questions to answer here pertain to machine interpretation—not just computer software and hardware, including scanners, but also microfilm readers and reader/printers. Which thresholds must be met for:

- Full-Text Searching: must text in the preservation copy be machine-readable? At what level of quality?
- Non-linear Navigation: does a preservation copy need to support functions such as ‘go-to-page,’ ‘advance *n* pages,’ or ‘go-to-section’?
- Aggregation of Reproductions or Component Parts Into New Contexts (e.g., new, dynamically-created objects): which functions, facilitated by which standards for mark-up, must be supported by preservation masters?

Sustainability Factors

After use, the second set of criteria for preservation copies pertain to longevity. Which principles should be codified to assure funders and other stakeholders that copies made today not only meet or exceed the usability of the original (perhaps compromised) printed material, but also meet or exceed its longevity (otherwise, why make preservation copies)?

Metrics in this category should enable practitioners to conform to minimal thresholds of best practice enabling collections managers to plan and manage the macro- and micro-level systems and services to enable both object and information persistence.

Specific arenas to mandate (or excuse) compliance include:

- Format Longevity: the Library of Congress’s Framework for Digital Formats (which can be extended to analog formats as well) is an ideal framework to facilitate discussions pertaining to format choices according to factors of sustainability, quality, and functionality.⁶ Again, it is important to distinguish the roles of preservation masters from preservation use copies. Use copies might be measured according to different expectations regarding lifespan. Where masters must presumably exist in physical, processible form for a defined number of years (e.g., ideally in centuries) to qualify as preservation copies, use copies may not only be ephemeral, but even virtual (i.e., generated dynamically by software upon demand; or manually in scan- or print-on-demand operations)
- Packaging Longevity: expressed and measured according to standards and metrics for stability of micro-environments and the related mechanisms to organize, label, protect, and retrieve objects
- Environmental Control: expressed and measured according to standards and metrics for stability of macro-environments, as well as defined criteria for security, monitoring, and emergency preparedness and response
- Replication: expressed and measured according to standards and metrics pertaining to the number of copies, and their location (both administrative and geographic)

Affordability Factors

Failure to meet minimal thresholds for usability or sustainability results in obsolescence. If one sanctions “preservation” strategies as those which mitigate obsolescence, then affordability must also be taken into account. The managed environments necessary to facilitate maximum life expectancies of *any* format are not inexpensive to build and maintain. Many preservation masters “live” in repositories that price services at annual rates. For all stakeholders in cultural heritage (not just the custodians of the printed originals and/or their preservation surrogates), it is relevant to consider, “What happens if custodians stop paying?”

Lack of capital hastens deterioration and obsolescence—perhaps just as dramatically as sudden changes in user preferences and expectations, or subtle-to-sudden fluctuations in environmental control.

There are four lifecycle costs that must be accounted for to sustain use of material for any designated lifespan. By minimizing costs in each area, greater periods of longevity can be purchased per dollar. Each lifecycle cost should be assessed according to the same question of compliance raised in previous sections. To be designated as preservation copies, must reproductions be assessed according to guidelines of affordability?

Production Costs (One-Time)

Assessments of reformatting strategies against criteria for use and sustainability cannot be made independent of cost. Does affordability effectively rule out certain strategies and rule in others—such as selecting 35mm sliver-halide film rather than 4" x 5" negatives to reproduce book pages—according to the best balance of quality (including use), longevity, and cost?

Technical specialists can make a useful contribution in this arena of developing a preservation framework by recommending the units at which costs should be measured and compared for reformatting multi-page objects: per character, component, page, section, volume, title?

Cost to Sustain (Annual)

As documented in a previous article, many variables contribute to the actual costs for sustaining a collection of preservation masters.⁷ What is notable about this arena of affordability is that costs are ongoing. Thus, once preservation copies have been made and are adequately managed, the costs to process masters and/or distribute use copies are moot if custodians cannot afford to pay the “maintenance” costs associated with managed storage, independent of rates of use.

As illustrated by the following example of 2,056 titles (at the time of writing this article) belonging to the Harvard University Library Open Collections Program’s *Women Working* collection, repository choices and format choices are likely to be meaningful factors when calculating maintenance costs. Like providers of commercial and non-profit repository (or “preservation”) storage, the Harvard University Library bills owners at annual rates, in order to recover some of the costs of managing the operations.

Harvard University Library Repository Storage Pricing

	annual \$	
Harvard Depository (HD) Film Vault (50°F/25% RH)	\$10.65	per Billable Square Foot*
HD Standard Vault (50°F/35% RH)	\$4.23	"
Harvard University Library Digital Repository Service	\$5.00	per Gigabyte

* 12" x 12" x 9"

Because Harvard bills at units of size, cost is relative to the number of preservation copies created and their size based upon their component parts and formats.

In the case of *Women Working*, the 2,056 titles now exist in three formats at Harvard: printed originals, and, as a result of the Open Collections Program digitization project, both 35mm preservation microfilm (two generations), and digital page images (produced either as compressed 1-bit or uncompressed 24-bit according to the meaningful content of each page).

The computed costs for storing all *Women Working* texts—with a mean average of 189 pages per title—in the best preservation environments offered by the Harvard University Library are as follows:

Annual Costs for Repository Storage: *Women Working*

	\$ per title	\$ collection
35mm microfilm (film vault)	\$0.21	\$ 436
Printed volumes (189 pp/vol)	\$0.22	\$ 456
Digital images (ave 687 KB)	\$0.62	\$1,273

Because microfilming significantly miniaturizes printed material, one would expect the cost of storing film masters to be less than print. This is true in the case of *Women Working*, but note that a full accounting of film storage costs takes into account the number of versions (2) and their location within the Harvard Depository. If, for example, Harvard’s owning libraries decided to store the second-generation print master film in the standard vault rather than the film vault, costs would be reduced by nearly 29% (to \$0.15 per title), but the film vault’s benefits of the lower RH and higher-performing air filtration would be sacrificed.

Similarly, the cost difference between digital and print, or digital and film is an attribute of the format choices made during digitization and the current rate of storage per GB in the Digital Repository Service. The costs to store digital images from another text digitization project could be significantly higher if all pages had been scanned in color. Annual storage costs would increase from \$0.62 to \$3.69 per title for compressed images (4.0 MB per page), and \$12.82 per title for uncompressed 300 dpi 24-bit images.

Cost to Process (Intermittent)

Until requested for use (by a computer or a person), or assessed as being at high risk of obsolescence, preservation copies do not need to be processed or delivered.

This point in the lifecycle of materials care is particularly relevant to evaluate the utility of the format selected (or being considered for) a preservation master against the expected (or known) delivery format requested by a user. Where 35mm microfilm, for example, might be judged as the most cost-effective and sustainable format for certain source materials (e.g., books without meaningful color), what might its processing costs be to generate the copy preferred by the user? Film is currently amenable to generating hard-copy prints (via reader/printers), service copy duplicate reels of microfilm (via film duplicators), or

digital reproductions (via film scanners and/or keying operators transcribing the source). If any of these intermediary processing technologies disappear, does microfilm effectively become obsolete because it is too difficult or too expensive to process? (The same assessment can be done for digital images—where today’s processing costs are sometimes misconstrued as being “free,” when managers fail to account for the operational costs of the infrastructure needed to generate digital use copies from the stored masters.)

Cost to Distribute (Intermittent)

Issues associated with costs to distribute are similar to, although technically separable from, costs to process preservation masters. As noted in the section on production costs, technical specialists can recommend the units at which costs should be measured and compared in these last two categories to meet distribution requirements associated with preservation management (e.g., managing multiple copies in multiple repositories) and user services.

Conclusion

Is microfilm a viable preservation technology for the next 20 years and beyond? Will the millions of volumes that Google digitizes be of preservation quality? Are there citable, consensus-supported guidelines, informed by standards, that the preservation community has adopted to sanction preservation? Because the answer to the third question is no, there is no definitive way to answer the first two questions.

Engaging technical experts and other preservation stakeholders in developing a usable framework to assess preservation strategies—beyond considerations and comparisons of formats and their purported attributes for longevity—will help to ensure that preservation strategies fulfill stated use and sustainability requirements at the lowest overall costs for creation, maintenance, processing, and distribution.

References

1. National Research Council, *Preservation of Historical Records*, National Academy Press, Washington, DC, 1986, pg. 6.
2. George Farr, Jr. and Deanna Marcum, *Preservation in the USA: A Case Study in Cooperation*, 67th IFLA Council and General Conference, August 16-25, 2001, available at <<http://www.ifla.org/IV/ifla67/papers/159-249e.pdf>>.
3. National Endowment for the Humanities, NEH Grant Management System Database, Division Director's Memos to the National Council Committee (cumulative), data provided by Charles Kolb, Senior Program Officer, Division of Preservation and Access, 2005.
4. Kathleen Arthur, et. al., *Recognizing Digitization as a Preservation Reformatting Method*, Association of Research Libraries, Washington, DC, 2004.
5. Nancy E. Elkington, ed., *RLG Preservation Microfilming Handbook*, Mountain View, CA, 1992.
6. Caroline R. Arms and Carl Fleischhauer, *Digital Formats for Library of Congress Collections: Factors to Consider When Choosing Digital Formats*, November 7, 2003 (DRAFT).
7. Stephen Chapman, *Counting the Costs of Digital Preservation: Is Repository Storage Affordable?*, *Journal of Digital Information*, 4, 2, (2003).

Biography

Stephen Chapman is Preservation Librarian for Digital Initiatives in the Weissman Preservation Center, Harvard University Library. He advises the Harvard community about approaches to collections digitization, and facilitates discussions and development projects to optimize systems and workflows for preservation digitization, digital preservation, and delivery of digital content. He recently published the Library Technology Report, *Techniques for Creating Sustainable Digital Collections*, and he co-authored the draft data dictionary for the NISO Z39.87 metadata standard.