Preliminary Results of Investigation to Optimize Microfilm for Digitization

This report summarizes Harvard's preliminary findings from its investigation into whether modifications to preservation microfilming procedures would yield better digital images in the film-first hybrid approach to reformatting brittle books. In this context, "better" refers to improving the quality or lowering the costs of black-and-white raster images (e.g., TIFF) created by scanning a 35mm print master. A related question of equal interest was, "How much would the cost of microfilming increase by instituting these enhanced-for-digitization modifications?"

The nature and timing of this investigation was occasioned by the shared commitment of the John G. Wolbach Library and the Harvard-Smithsonian Center for Astrophysics to preserve and improve access to historical astronomical literature. (The Smithsonian/ NASA Astrophysics Data System was an operational digital library at the outset of the NEH microfilming project, and the research community it serves has voiced a strong desire to access the entire corpus of astronomical literature in digital form.) In order not to disrupt the schedule of the NEH microfilming project, we focused on the serial literature in astronomy as a manageable subset to "enhance" for digitization. In this manner, our investigation encompassed 454 volumes of 21 titles, for a total of approximately 212,000 pages (131 reels of microfilm).

These findings are reported as preliminary since the enhanced reels of microfilm (hereafter referred to as "prototype microfilm") have not yet been scanned. Since completing microfilming, we have distributed a Request for Proposal for microfilm scanning and have reviewed several responses (which include sample 600 dpi 1-bit digital images). When we complete this hybrid reformatting project for historical astronomical literature, we will publish an article detailing the full project methodology, tools and techniques, costs, and findings.

Modifications to Microfilming Workflow

Based upon the findings of Yale University Library's *Project Open Book* and, more importantly, the willingness of the staff in Harvard College Library Preservation & Imaging Services to accommodate procedural changes in microfilming, we instituted a number of modifications to microfilm preparation, filming, and film inspection:

Table 1. Optimize-for-Digitization Microfilming Procedures, and their rationale*

procedure		quality	cost
prep	collate volume (transcribe pagination and designated features) to pre-create structural metadata for digital object; print "Collation Target"		X
	reel programming: make sure that largest volume is filmed first	X	X
filming	in addition to microfilm resolution chart, film targets appropriate to evaluate film scanners (RIT Alphanumeric Test Object, Kodak grayscale)	X	
	film Collation Target		X
	film *every* page, including consecutive blanks (to be consistent with record of collation and to facilitate future production of digital codex)	X	
	center all frames on reel and minimize skew to greatest possible extent	X	X
	create additional print master (to allow for possibility that film will be scratched or otherwise compromised by high-speed scanning)		X
	blip frames that contain full-page illustrations	X	X
inspection	carefully note all second exposures and number of frames for foldouts		X

* expressed as desire to maximize digital quality ("quality") or to minimize digital cost ("cost")

Findings and Recommendations

We accounted only for the two procedures that yielded "meaningful" increased costs to create the prototype microfilm. Creating the second print master added a cost of $13.00 per reel, and as noted in Table 2 below, creating structural metadata added approximately $10.00 per volume. The mandatory inclusion of scanning targets added two frames per reel. The Collation Targets also increased the frame count by one to two frames per title. (Costs were not computed for these additional frames.) HCL Preservation & Imaging Services reported that, after appropriate training, technicians produced the blipped, centered-on-reel prototype microfilm at the same rate as "standard" preservation microfilm. Again, there was minimal impact to the time spent on microfilm inspection since HCL Preservation & Imaging Services routinely produces a written "Roll Report, Negative Microfilm" that accounts for second exposures and other anomalies.

Structural Metadata. The premise of creating structural metadata from paper — which increased the time of microfilm prep by an average of 20 minutes per volume — was that the added investment in the microfilm would be more than recovered during digitization, particularly if scanning were deferred for a significant period after microfilming. As Table 2 shows, however, gathering comparable amounts of metadata from digital images adds only a nominal cost per volume (less than $1.00 at Harvard's 1999 wages). Moreover, evidence is growing that there are two significant benefits to moving metadata creation to the end of the (film-first) hybrid workflow: (1) metadata is more accurate; we have already found discrepancies between the metadata gathered from paper and the *actual* number of images produced from the microfilm (suggesting that either errors were made during collation or that collation and filming were out of sync); and (2) creating metadata after scanning facilitates 100% inspection of the digital images.

Table 2. Comparison of Times & Costs to Create Structural Metadata from Paper, Microfilm*, and Digital Images for a 300-page brittle book (monograph or serial)**

source	point in film-first hybrid workflow	time	cost+
original printed vol.	beginning (microfilm prep)	.33 hours (20 mins.)	$6.66
microfilm	middle (after filming, scanning prep)	.50 hours (30 mins.)	$7.66
digital images	end (after filming, after scanning)	.39 hours (24 mins.)	$7.36

* findings from Yale University's Project Open Book; see. Fig. 11, "Processing Costs," in Paul Conway, "Yale University Library's Project Open Book: Preliminary Research Findings," D-Lib Magazine, February 1996, http://www.dlib.org/dlib/february96/yale/02conway.html.

** findings from Harvard University Library Digital Initiative project, Online Historical Reference Shelf, to digitize 103,000 pages of annual reports from Harvard University and Radcliffe College, 1825-1995

+ costs are actual rather than normalized costs of the wage (salary + benefits) paid to the technician in the year in which the work was peformed (1998-99 wages at Harvard, 1995 at Yale)

Having sent duplicate print masters of a total of six reels of prototype microfilm to three microfilm scanning vendors (as part of the RFP for scanning), we are encouraged by the results. Contrary to our assumptions, lightbox inspection of the scanned film revealed no visible scratches or other defects introduced by the microfilm scanner or scanning technician. (We should note, however, that these results are associated *only* with the SunRise Imaging Proscan microfilm scanners.) Thus, the additional cost of producing, and maintaining, a second print master *may not* be justified. Further testing is recommended.

Since careful attention to the consistent placement of images on film appears to add no increase to costs, we recommend adopting these practices (to minimize skew and center volumes per frame). Consistent image placement not only minimizes the need to deskew or crop images scanned from microfilm, but it also produces images on film that will be consistently placed in microfilm readers – and thus easier to read.

Findings and recommendations regarding the benefit of blipping microfilm — either to identify frames for grayscale scanning or frames for indexing, or both — will be reported after we have assessed the results of production scanning.

Conclusion

We enter the final phase of the film-first hybrid investigation with several premises: that creating structural metadata from print, although cost effective, is prone to error; that attention to placement of images on film, as well as the inclusion of technical targets at the beginning of each reel, will yield higher digital image quality; and that the information recorded in the Roll Reports during microfilm inspection will be invaluable in helping to check-in, inspect, and otherwise process digital images and their associated structural metadata. We will gather statistics in each of these areas and report our findings in the paper we will submit for publication.

Available from **http://preserve.harvard.edu**
http://preserve.harvard.edu/pubs/microfilm_for_digitization.pdf
Last updated: September, 2000

