Selection for Preservation: A Digital Solution for Illustrated Texts

Janet Gertz

The point of selection for preservation is to identify endangered library and archival materials that have long-term intellectual value and are therefore worth the effort and cost of long-term preservation. Technical issues cannot be separated from selection for preservation, because the limits of preservation technologies can influence decisions. At least for the present, we need to combine digitization with analog preservation methods. This "hybrid approach" operates on an assumption that we can in fact make a digital version of the original, and that the digital version will be able to serve the needs that justified selecting the item for preservation in the first place. Columbia University Libraries' Preservation Division has been experimenting with the hybrid digital approach, selecting digitization as the preservation method for materials that previously had to be rejected because our reformatting technologies could not copy them in a way that made the contents accessible to users. In 1994 Columbia undertook a project funded by the Commission on Preservation and Access to combine film with digitization and test the hybrid approach on illustrated materials. We have demonstrated that scanning the microfiche can, in fact, produce digital images with legibility equal to the images made directly from the original printed maps. While legibility was quite successfully achieved during the project, questions remain about the quality of the color that can be delivered to the viewer. Capture is one side of the coin, delivery is the other.

The point of selection for preservation is to identify endangered library and archival materials that have long-term intellectual value and are therefore worth the effort and cost of long-term preservation. The goal of preservation reformatting is to produce a copy of a brittle or damaged original that captures as much as possible of the intellectual content. That copy must be long-lasting and at the same time fully usable by scholars. A long-lasting copy that cannot be used is not much improvement over the brittle original. This includes digital images. The potential for enhanced uses of digitized materials is almost unlimited. But if the digital image

JANET GERTZ is Director for Preservation, Columbia University Libraries, New York (gertz@columbia.edu). This article is based on a presentation to a joint program meeting of the Collection Management and Development Section and Preservation and Reformatting Section, Association for Library Collections & Technical Services, at the ALA Annual Conference, Chicago, Illinois, June 26, 1995. Manuscript received September 1, 1995; accepted for publication October 17, 1995.

is illegible or badly indexed, then preservation and access both fail.

Determining what is usable rests on the selector's analysis of research patterns in the disciplines that use the materials. How will the materials be used—for a careful reading of the full text, for quick reference consultation, to collect data, or to be examined in close combination with other items, as when an art historian compares tens or hundreds of images? Will use be frequent or infrequent over time? Heavy in the short term? Low in the long term? Must the materials be immediately available or can scholars tolerate a slower retrieval rate?

Technical issues cannot be separated from selection for preservation, because the limits of preservation technologies can influence decisions. Is there a way to return the item itself to continued use? Is there a workable reformatting method for replacing the item with a copy that can serve the same uses as the original? How much of the information contained in the original can be transferred to the copy? If we are talking about reformatting via microfilm, is the item one whose content can be captured and reasonably used on film? If we are talking about reformatting through digital imaging, the same questions apply: Can the item be scanned successfully? Can we achieve image quality that will serve the desired uses? Funds are always limited and neither digital nor analog preservation methods are cheap. It is important to choose the preservation and access methods that can really serve scholarship.

Digital Preservation Technology

Digital technology offers us the potential to broaden preservation because of its ability to enhance access to endangered materials. But at the same time the validity of digitization as a source of long-term preservation is very much an open question. We know that digital storage media have a short life relative to microfilm, or even relative to acid paper. We know that software and hardware change with almost frightening speed, leaving older iterations behind and often unreadable. It is too soon to rest secure that repeated refreshment and migration will not somehow alter the content of the digital files.

At least for the present, therefore, we need to combine digitization with analog preservation methods. For true long-term security, we must assure that we have the most permanent analog version we can achieve—this might be the original item itself, properly repaired and housed, or a reformatted version on film or other stable medium. Microfilm in long-term archival storage can continue to serve for centuries. After all, longevity and stability are the hallmarks of properly made and stored film.

In parallel to this archival-quality analog version, we might also want to make what could be called a "digital preservation" version. In doing so, the original is scanned at the highest resolution needed to produce full legibility; gray scale or color are used when the original requires it, and are carefully matched to the original; and the master copy is maintained in a lossless format. The digital preservation version presents an accurate record of the original for scholarly purposes, without further enhancement. Certainly digitization offers us tools to enhance images; a manuscript with text obscured by coffee stains can be digitally altered to be much more legible, but this changes the facts of what the original really was at the time of scanning. Authenticity and accuracy in representing the original are particularly at issue where the original will be discarded after scanning.

Producing Virtual Copies from Digital Versions

From the digital preservation version scholars can then derive as many copies as they wish, and it is these use copies that can be enhanced and manipulated at will. The point is both to facilitate multiple possibilities of use without compromising the authenticity of the digital preservation version, and to maintain the analog version for those who will need to consult it and, of course, in case of accidental loss or change to the digital preservation version.

Should it ever be needed or desired, we will have the analog version to rescan.

This model, which has been called the "hybrid approach" (Willis 1992), assumes that we can in fact make a digital version of the original, and that the digital version will be able to serve the needs that justified selecting the item for preservation in the first place. In trying to decide whether digitization is appropriate in preserving a specific item, it seems to me that we must go back to the basic question: Can the applications and type of access required by users best be served by paper, film, or an online version? How might scholars use a digital version? Would it provide anything the analog version cannot?

If we think digital technology is the answer, then we must determine what levels of resolution and accessing mechanisms are appropriate. If we want full access online to high resolution images to serve as surrogates for rare or fragile materials, or as replacements of originals to be discarded after scanning, the question is whether current technology is capable of capturing, transmitting, and displaying the needed information.

Selection decisions must be made with an understanding of where we run up against the limitations of the technology, and where we reach the point that digitization offers little or no improvement over analog preservation, or does so only at great cost. Determining how selection for preservation will work in a digital environment is a moving target because technology changes so rapidly. We can talk about selection for what is technically possible now, but we must also identify how we want to use digitization in the future, and then push the technology in those directions. As what is available on the technological side continues to change, we will broaden the range of materials appropriate to select for preservation through digitization.

Case Study: Color Oversize Maps

Columbia University Libraries' Preservation Division has been experimenting with the hybrid digital approach, selecting digitization as the preservation method for materials that previously had to be rejected because our reformatting technologies could not copy them in a way that made the contents accessible to users.

Background: Preservation Reformatting Methods

Traditional preservation reformatting methods can lead to negative selection when we decide not to select an item because we have no satisfactory preservation and access method. Illustrated materials characterized by color and large size are prime examples, especially art and architecture publications and those of geology and geography, with their maps and charts. These genres present a particular preservation challenge. It is essential that scholars be able to view illustrations while they read the accompanying text. Further, they need to see the illustration as a whole, to follow information across the breadth of its surface, and they must also be able to read the finest details at every point. Color and pattern are important for aesthetic reasons and as coding devices on maps and charts (Commission on Preservation and Access 1989; Joint Task Force 1992).

Preserving illustrated materials is not easy. Oversize illustrations in brittle volumes suffer particular physical stress. Publishers combine text with illustrations by folding oversize items into the binding or pockets. While the pages of the volumes follow the usual course of slow chemical degradation, self-destruction of the oversize foldouts moves swiftly because the brittle folded edges break off and information is irretrievably lost.

This is not an insignificant problem. In a random sample of brittle architecture monographs, the Preservation Division found that 68% of the collection contains a mixture of illustrations not suitable for filming; 17% have oversize foldouts. Similarly, the University of Chicago Preservation Department reports that between 10% and 20% of their collections in the sciences contain oversize graphic materials (Preservation Department 1989).

Black-and-white microfilm certainly is not a successful method for preserving print materials that are heavily illustrated with color and oversize elements. Not only is the color lost; the oversize illustrations must be filmed in sections in order to keep them legible, and the result is a major loss of functionality. Sectioned illustrations are often genuinely unusable.

Even torn and crumbling original illustrations can be preferable to poor reproductions. At Columbia, art and architecture faculty rejected microfilm of illustrated originals. Because microfilming did not preserve the content of the volumes and because hand conservation was not feasible, selection for preservation of brittle illustrated books in essence became selection for indefinite residence in the "sick bay" instead of selection to reformat the books and return their content to scholars for research.

Finding a new way to preserve text plus illustration thus ranks very high among Columbia's preservation priorities. Collection development specialists are extremely conscious of the collection's preservation needs (many of the older materials are brittle), but are also extremely frustrated by the technical limitations of preservation that have prevented them from making successful preservation decisions on these very important conspectus level 4 and 5 collections, which would otherwise be among the first selected for preservation.

Methodology

In 1994 Columbia undertook a project funded by the Commission on Preservation and Access to combine film with digitization and test the hybrid approach on illustrated materials. An important question for the project was whether scanning a film intermediary would produce results comparable to scanning the original. The other area for investigation was to experiment with a means for integrating the digital files of the text and illustrations into an online whole. The goal was for the scholar to view the text and the illustrations in juxtaposition, as was possible with the original paper volume.

In phase one of the project, we addressed the scanning of oversize color illustrations, specifically maps. Capture was equally successful from film or from the original. We also have found, not surprisingly, that Internet access and delivery have definite limits at present and that printouts can serve as a stopgap in the interim. A detailed final report of the project and over three hundred digital images can be accessed over the Internet at http:// www.cc.columbia.edu/imaging/html/largemaps/.

Moving through film to digital was our preferred option for a number of reasons. For instance, several of the vendors working on the project prefer scanning a film intermediary rather than the original, claiming that they can achieve a better scan from the film version. In practical terms, when the original is very fragile, handling needs to be minimized. Many scanners cannot accommodate oversize materials, whether fragile or not. There is also the fact that many scholars are still limited in their ability to access digital files (especially the very large files needed for pictorial materials) and might still choose to access the volumes in the film version.

During the project we worked with five turn-of-the-century brittle maps from the *New York State Museum Bulletin*. We compared digital images made directly from the paper originals with digital images made from single-frame color microfiche of those same maps. Single-frame microfiche use the entire field (normally 105 x 145 mm) to carry just one image. The microfiche had been produced during an earlier Commission on Preservation and Access project, which proved that such microfiche could successfully capture and preserve an oversize illustration's content in fine detail at a low reduction ratio (Klimley 1993).

We have demonstrated that scanning the microfiche can, in fact, produce digital images with legibility equal to the images made directly from the original printed maps. The smallest print on the original maps is one millimeter high and can be read equally well in the online versions produced by scanning the microfiche and in the versions made by scanning the origi-

nal maps. When scanned at a pixel depth of 24-bit color, a resolution level of 200 dots per inch (dpi) on the original map produced full legibility of the smallest type.

We are used to hearing that 600 dpi is needed for preservation, but that is in the context of black-and- white (binary) scanning. The use of 24-bit color adds a great deal of visual information to the image. Higher pixel depth (that is, use of gray scale or color rather than black and white) allows use of lower resolution, so that 200 dpi in 24-bit color gives legibility approximately equivalent to 600 dpi binary (Ester 1991). What this means for oversize images is that a map twenty inches across requires 4,000 dots (20 inches x 200 dots) across its surface in order to reproduce the finest one millimeter print legibly in full color. The microfiche of that same map also needs 4,000 dots across the surface of the map image to capture the same degree of detail as the original. On the microfiche the map image is perhaps only four inches wide, so that what is needed is four inches at 1,000 dpi when scanning the microfiche in order to reach 4,000 dots across the map surface.

While achieving legibility was quite successful during the project, questions remain about the quality of the color that can be delivered to the viewer. We can capture color with 24 bits of information per dot, which translates to a potential for sixteen million different hues. Capture can be very accurate if scanners are carefully calibrated using standard color charts, but it is also true that certain scanners are biased toward certain color ranges, just as some films are "cooler" in tone and others are "warmer."

Monitors normally display only 256 colors. No two printers or monitors can be guaranteed to output exactly the same shades unless they have also been carefully calibrated. All of this means that the color we see online is not terribly true to the original maps. However, in the case of maps, color is primarily used for coding, so that most scholars are satisfied as long as all the codes remain distinct and the color approximates the original. But scholars who need full color accuracy (for instance, art historians) might well find digital copies less than satisfactory. The degree to which the color of the film intermediary does or does not match the paper original is also relevant.

Conclusions

Our conclusions, then, hold only for genres like modern printed maps, where the information is partly textual (place names, labels, numbers), partly linear (roads, borders), and partly codes made up of a limited number of colors and patterns. There is a definable resolution at which we can say all information has been captured and is fully legible, just as we can for a printed page. There is no particular gain in using yet higher resolution because there is nothing further to capture. This contrasts with the situation in scanning works of art or historical artifacts with many subtle color tones, and where important information content may be contained in the very fibers of the paper.

Capture is one side of the coin, delivery is the other. We can currently capture more information than can readily be transmitted or displayed on an average monitor. The files of the scanned maps at high resolution and 24-bit color can run as large as twenty megabytes when uncompressed. What Columbia has mounted on its Web site for Internet access are lossless GIF versions of the files with 256 colors and lossy JPEG versions with 16 million colors. The resolution is cut back to about 150 dpi. These files run up to about six megabytes. Unfortunately, these lower-resolution files that are more easily transmitted and viewed carry too little detail for the largest maps to be fully legible online. We can hope that this situation is temporary and that it will be corrected as hardware and Internet delivery times improve. Meanwhile, we also produced full-size, fully legible paper printouts from the high-resolution files as a use medium.

Phase two of the project is now underway, with a goal of reassembling the text and illustrations online. We will scan the microfilm of the text of four volumes of the *Museum Bulletin* at 600 dpi in black and white, and will scan single-frame color microfiche of the illustrations in 24-

bit color at 200 dpi. This will produce a series of digital files, one for each page and one for each illustration. We will use indexing and document structure software to integrate the files of pages and illustrations so that users can move easily from one to another online. The result will be the full preservation of the four sample volumes. Long-lasting microfilm and microfiche will be created for all the text and illustrations, along with a digital version that maintains the author's juxtaposition of words and illustrations online, and through paper printouts.

Obviously, many questions remain. To what extent will the existence of these digital images satisfy the scholarly community's needs? Will they affect how scholars do their research? What role will they play for scholars interested in detailed analysis of the maps? Will they serve as pointers to the originals that must then be consulted, or to requests for printouts, or will some scholars be able to do their work with the digital images alone? Will the quality of the color images be satisfactory? Will a bit-mapped version of the text suffice?

Finally, what does all of this mean for selection for preservation? It appears to offer potential for a new option for selectors faced with trying to preserve one very difficult class of materials. And it offers one criterion to justify selecting digitization as a preservation method: to employ digital imaging on the grounds that a digital version can solve preservation problems that cannot be handled through analog means. Digitization broadens our ability to capture information from a wider range of media, so that more preservation decisions can be made based on the condition and content of the item instead of on the limited range of traditional technical options for preservation.

Works Cited

Commission on Preservation and Access. 1989. *Scholarly resources in art history: Issues in preservation. Report of the Seminar, Spring Hill, Wayzata, Minnesota, September 29–October 1, 1988.* Washington, D.C.: Commission on Preservation and Access.

Ester, Michael. 1991. Image quality and viewer perception. *Visual resources* 7, no. 4: 51–63.

Joint Task Force on Text and Image. 1992. *Preserving the illustrated text.* Washington, D.C.: Commission on Preservation and Access.

Klimley, Susan. 1993. Notes from the cutting edge. *Microform review* 22: 105–7.

Preservation Department. 1989. *Report on the condition of the collections.* Chicago: University of Chicago Library.

Willis, Don. 1992. *A hybrid systems approach to preservation of printed materials.* Washington, D.C.: Commission on Preservation and Access.