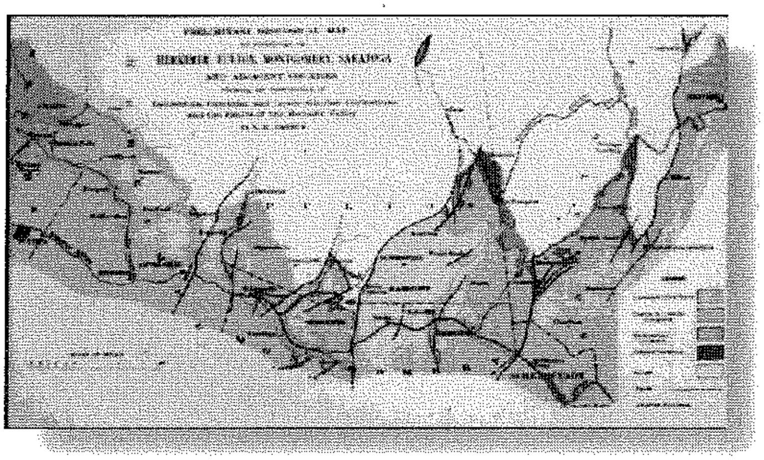
The Commission on Preservation & Access

Oversize Color Images Project, 1994–1995

A Report to the Commission on Preservation and Access



August 1995

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

Oversize Color Images Project, 1994–1995
Final Report of Phase I

A Report to the Commission on Preservation and Access

by Janet Gertz
Director for Preservation, Columbia University

Project Management Team
Janet Gertz
Robert Cartolano (User Services Lead Consultant,
Columbia Academic Information Systems)
Susan Klimley (Geology Librarian)

August 1995

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2217

This report describes a seven-month investigative study conducted under contract to the Commission on Preservation and Access to assess the technological possibilities for reformatting brittle maps. The project, part of the collaborative activities of the Digital Preservation Consortium, was conducted by Columbia University Libraries and Academic Information Systems. The study sought to identify the most acceptable preservation and access techniques available for oversize, color images associated with text. Results will be used to develop plans for a large-scale project to scan and preserve the publications of the United States Geological Survey.

Additional copies are available for $10.00 from The Commission on Preservation and Access.
Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access," with payment in U.S. funds.

This paper has been submitted to the ERIC Clearinghouse on Information Resources.

∞ The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1992.

ISBN 1-887334-42-4

TABLE OF CONTENTS

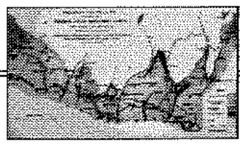
Introduction

This report contains the results of Phase I of the Oversize Color Images Project to identify acceptable preservation and digital access techniques for dealing with oversize, color images associated with text. The goal of Phase I was to provide a preservation-quality photographic archival copy, a digital version for on-line access, and paper printouts. Five maps from brittle volumes were scanned, along with single-frame color microfiche of the maps (produced during an earlier contract with the Commission on Preservation and Access) and 4"x5" color transparencies. Printouts were made from many of the digital versions.

Image quality, in terms both of capture and of display, involves a number of complex factors that are not always easy to consider and evaluate separately: characteristics of the original media, characteristics of the photographic intermediaries and how they were produced, scanning and display resolution, compression, storage, display hardware, display software, and network transmission limitations. This report concentrates primarily on resolution and on display issues relevant to modern printed maps, which consist of text superimposed on images made up of a relatively small range of colors without subtleties of shading.

Evaluation of the digital files and the highest quality printouts shows that fine details can be captured successfully from all three media (paper original, microfiche, transparency). For the purposes of creating an archival-quality storage medium combined with the potential for increasingly sophisticated electronic presentation of such materials, the project has been very successful.

At this time the ability to capture information outstrips capacity for easy access and display with average equipment, so that immediate on-line use of high resolution files is somewhat limited. Because all printed details visible in the paper originals can be fully captured by current technology, however, rescanning should not be needed in future even though access and display capacity is expected to increase rapidly in the near future.



Rationale for the Project: Problems of Preservation and Access Presented by Oversize Color Materials

The Commission on Preservation and Access paper *Scholarly Resources in Art History*[1] and the report of its Joint Task Force on Text and Image, *Preserving the Illustrated Text*,[2] have documented the inadequacy of traditional microform techniques to preserve brittle books characterized by a complex combination of text and illustrated materials. Such materials are relatively prevalent especially in the literature of certain disciplines. The University of Chicago Library's Preservation Department's 1989 "Report on the Condition of the Collections"[3] indicates that 10-20% of their collections of general science, astronomy, natural science, agriculture, technology, and history of science contain oversize materials folded into bound volumes. A survey in 1994 by Columbia University Libraries Preservation Division of architectural history volumes found that 68% contain halftone, continuous tone, or color illustrations; 17% contain oversize foldouts. Geography and the geologic sciences are similarly characterized by heavy use of maps and other oversize illustrations.

Brittle volumes with oversize images suffer particular physical stress because the juxtaposition of text and images was achieved through folding oversize items into the binding or pockets. Although the pages of the volumes follow the usual slow course of chemical degradation, self-destruction of the oversize foldouts is dramatically hastened as every folded edge breaks and information is irretrievably lost.

Maps present a particular challenge. Scholars need to be able to see the map as a whole, to follow information across its surface, and to read fine details at every point. They have frequent need for paper copies to carry away from the desk. Color and pattern are both important as coding devices. Any reproduction must distinguish among colors that are distinct in the original and distinguish among patches of the same color printed in distinct patterns, such as parallel horizontal or vertical lines.

Although high contrast black and white microfilm serves well as an archival medium, it is inadequate to (1) convey color, (2) capture fine details of large images, (3) maintain the juxtaposition of image and text, and (4) provide methods for printing out full-size color copies of the images. When a map is filmed as a whole, the details are lost because the reduction ratio must be very high in order to fit the map into a 35mm frame. In order to capture the details, a map must be filmed in sections that can only be viewed sequentially. Printouts are often illegible and in any case must be pieced together to recreate the entire map.

Concern over these issues resulted in the Commission's 1991-1992 contract to experiment with preserving twenty-eight brittle numbers of the *New York State Museum Bulletin* through creation of single-frame color microfiche[4] and black-and-white scanned images.[5] Paper printouts were made from both versions. The sampled numbers, in subjects ranging from botany to entomology to the earth sciences, contained a wide variety of problematic and often oversize illustrative material including color drawings and maps. Although much was learned from that experiment of two years ago, one of the most tenacious problems proved to be the oversize color images.

Identification of a group of suitable preservation and access mechanisms for oversize images has remained an essential first step in preserving the literatures that utilize such images. Digitization offers a wide new range of access options. A bitmapped image of a map captured at high resolution and with 24-bit color is a detailed copy of the map that can be saved in lossless Tagged Image File Format (TIFF format). A wide variety of use products can then be derived using Joint Photographic Expert Group (JPEG) compression or other space saving but lossy compression methods.

At the same time, the desire for long-term preservation of the information coupled with uncertainty about long-term digital refreshment and migration mean that we continue to want both a film master copy and a digital version. Most of these books and their oversize illustrations are brittle to the point of crumbling so that keeping the originals is not enough; yet, it is still not possible to feel confident enough to rely solely on the digital copy for long-term preservation.

Goals of the Project

The goal of the present project was to devise a methodology for providing both a film version of the image to serve as the permanent archival copy and a digital version to serve as the access copy. Whether by microfilming or scanning, much of the cost of reformatting lies in selection, cataloging, and preparation of the items and in actually handling the item during capture. Ideally, the information should be captured directly from the original image only once and then transferred to other formats as desired. Since in this case the originals are not only oversize but also brittle, it would put less strain on the fragile originals if a film version could be made first and then scanned.

The questions addressed in this project include:

- How do digital versions made from the film copy compare to digital versions obtained by scanning the original?
- What level of resolution is needed to capture at least as much information (and preferably more) as the traditional methodologies?
- Will users accept the digital version and be able to use it to their own satisfaction?

To answer these questions, the project compared the results of scanning five oversize, color, brittle maps using a variety of methodologies to produce both onscreen versions and paper printouts. Geologists, geoscientists, other subject specialists, librarians, and other interested parties offered reactions to the digital and paper versions.

Description of Maps to be Digitized

Five oversize, color, brittle maps from the *New York State Museum Bulletin* were selected for the project and borrowed from the New York State Library. These are the same copies of the maps that were used in the 1992 project described above. The size of the maps varies from 9"x12" to 19"x24"; thus, while they are larger than normal page size, they are actually relatively small for maps.

The project worked from three formats and compared the results of scanning each version.

Original paper maps

1. "Geological Map of Perce and Vicinity"; NYS Museum Bulletin 80 (1905); plate; 9"x12.5"; Hand-drawn cross-section with watercolor-like pastel pigmentation.

2. "Part of the Catskill Quadrangle"; NYS Museum Bulletin 84 (1905); plate 7; 14.75"x9"; Three-color geologic map with contour lines in light brown, water in blue, and details in shades of pink.

3. "Preliminary Geological Map of Portions of Herkimer... and Adjacent Counties"; NYS Museum Bulletin 95 (1905); plate 14; 8.75"x14.5"; Geologic map with coding in one color (light brown) in a variety of patterns and intensities.

4. "Surface Geology of Mooers Quadrangle"; NYS Museum Bulletin 84 (1905); loose map; 24"x17.5"; Typical full-color geologic map with fine detail and subtle color coding.

5. "Geological Map of Buffalo Quadrangle"; NYS Museum Bulletin 99 (1906); loose map; 21.5"x19"; Typical full-color geologic map with fine detail including details of Buffalo city proper and cross-section printed perpendicular to map along right edge.

The smallest type present on each map is 1 mm. Often this is the contour line elevation, which in the most challenging cases is printed in light brown in italics on top of color-coded areas. The result is very little contrast between the number and the background, and the numbers are difficult to read even in the paper originals. Even smaller are .6 mm squares representing buildings on the two larger maps, although these are in black and therefore easier to see.

The number of colors and their use and distribution on the maps varies. The simplest map uses only one color, printed in five codes that differ in pattern and intensity. The most complex map uses each of five distinct colors in three to four different patterns and intensities, for a total of seventeen different codes. The differences between the codes are often subtle, and it can require close examination of the original to distinguish among them.

Single-frame color microfiche

The New York State Library's copies of the five maps had been used in the 1991-1992 project by the German company Hermann and Kramer to produce single-frame color 105 mm microfiche on Cibachrome stock. A second generation set of those microfiche belongs to Columbia University and was used in the current project. The microfiche are 5.75"x4.13"; the actual map images ranged from 3.25"x3.38" to 3"x4.88". Three vendors scanned these fiche.

One vendor produced its own set of single-frame microfiche from the same paper originals. Their fiche are also on Cibachrome stock but are first generation. The size of the images differs. The maps of Perce, Catskill, and Herkimer range from 2.5"x1.75" to 2.88"x1.5", while the Mooers and Buffalo maps are approximately 4.75"x3.5".

Transparencies

Color transparencies (4"x5") were made from the same paper maps in the Columbia University Libraries Reprography Laboratory for the purposes of this project. Standard color charts were included with the maps. The film stock was Ektachrome 100 Plus.

Selection of Vendors

A request for proposal (RFP) was prepared and sent to a number of potential vendors in order to determine which would be interested in participating in the project and to see what imaging services might be available. The choice of vendors to whom the RFP was sent was based on previous contacts through other projects, the vendor fair held during the Research Libraries Group (RLG) Digital Imaging Technology for Preservation Symposium (March 17-18, 1994), and recommendations from colleagues.

The RFP sought vendors capable of:

- scanning directly from the oversize brittle originals by methods that would not damage them (only scanners where the originals would not be subject to motion were accept-

able for scanning because of the high potential for damage during movement);
- scanning from single-frame microfiche;
- scanning from 4"x5" transparencies;
- producing full-size, full-color paper printouts from all of the digital versions.

Vendors were required to:

- work with originals as large as 20"x30";
- provide data on CD, 8mm tape-TAR format, or QIC150 tape-TAR format;
- use any of the following file formats: JPEG, Kodak PhotoCD or PhotoCD Pro, TIFF, GIF (Graphic Image File format);
- provide minimum resolution of 2,000 x 2,000 pixels;
- capture color at 24-bit, 16 million colors;
- produce full color printouts, as close as possible to full size of the original.

Eventually, five vendors responded positively to the RFP and took part in the project.[6] The maps, microfiche, and transparencies were sent to each vendor in turn. No one vendor was able to offer all of the services listed above, and each used somewhat different scanning and printing methodologies.

In scanning the originals, the vendors were asked to include standard color charts to allow for evaluation of the color quality displayed at the screen and printed out in paper versions. The second generation microfiche (produced during the 1991-1992 project) unfortunately had not been shot with color charts. See *Description of Vendor Products* in the **Appendix**.

Definitions for Evaluating Quality of Project Images

When evaluating images, how does one define success? The purpose of this project was not to achieve the best possible quality image but, rather, the image that is good enough to serve fully the needs of researchers in lieu of the original paper maps. In order to do this, a reproduction of the original map must reproduce all details and color codes accurately. Quality control was performed on a Tektronix 19" 1280x1024 X terminal to ensure consistency, but images also were viewed on a variety of other equipment.

The definition of adequate quality is:

- *legibility:* All of the print can be read on the screen and on the printouts, including the 1 mm contour elevations, although the print may be somewhat fuzzy.
- *color accuracy:* All color codes are distinct on the screen and in printouts, even if colors have shifted in comparison to the printed originals.

The definition of fully successful quality is:

- *legibility:* All print, including the 1 mm contour elevations, can be read on the screen and on the printouts with full clarity.
- *color accuracy:* All color codes are distinct on the screen and in printouts, and there is no color shift in comparison to the printed originals.

Project Results

In looking at the products delivered by the vendors, it is clear that adequate results starting from the original map, the microfiche, or the transparency are achievable. The 1 mm elevation marks are legible at high resolution and all color pattern codes can be distinguished. There has been some color shift in all printouts and in all images as viewed at a variety of monitors, however.

These results mean that for preservation purposes it is safe to continue to create preservation-quality microfiche and transparencies to serve as surrogates or replacements for brittle originals. The film intermediaries can be later scanned with confidence even if the originals have been lost, and they will still capture the intellectual content of the originals.

It must be emphasized that only modern printed maps have been dealt with here, and that the goal has been to capture the printed content; in these materials this consists of text superimposed on a relatively small range of colors without subtleties of shading. Quite different results would be expected for archival materials, reproductions of works of art, and other images with a wider palette and more subtle detail.

It should also be noted that some vendors were more successful than others at producing digital images that met the definitions of adequate quality. For instance, although most of the vendors provided digital images with resolution of at least 150 dots per inch (dpi) or higher, in some cases the images were in better focus than others. Variations can be due to the particular scanners used, operator skill, and similar factors. In the same way, some vendors' images have better color quality than others', thanks to a combination of quality of lighting and care in balancing color and calibrating equipment.

Effective Resolution

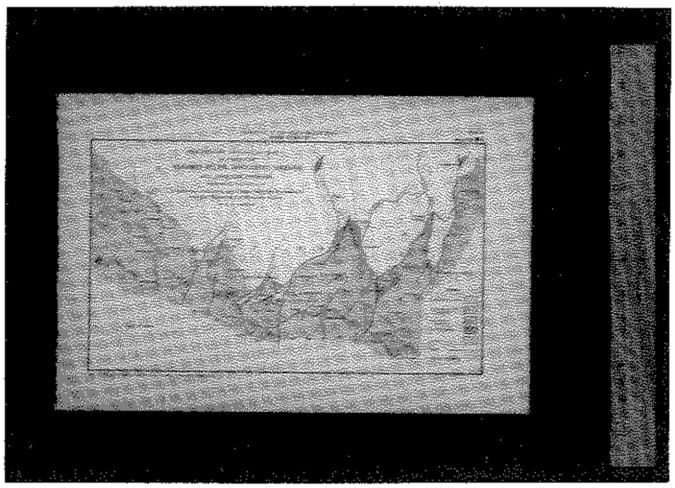
Effective resolution refers to how many dots per inch the digital file provides relative to the paper original, regardless of whether the image was made from the paper original, the microfiche, or the transparencies. Thus, if an image consists of 2,000 x 2,000 pixels, and it represents a scan of a 10"x10" original, the effective resolution is 200 dpi (2,000 pixels divided by 10 inches). Calculation of effective dpi includes the space devoted to color bars. If the same 10" original were accompanied by a color bar that is 3" thick and placed 1" from the map, the width would be 14" and the dpi would be 143 (2,000 pixels divided over 14 inches).

When discussing the resolution needed to capture the same degree of detail from a film intermediary as from the original, effective resolution is determined relative to the original: The film version must be scanned so as to provide the same number of pixels across the face of the miniature image as were used in scanning the full-size original. Assuming that the 10" original printed map has been filmed as a 4" image on a microfiche, each inch of the microfiche represents 2.5 inches of the original. Since the 10" original requires 2,000 pixels to achieve an effective resolution of 200 dpi, the microfiche of that map will also require 2,000 pixels. But the 2,000 pixels which stretched across 10" of the paper map must now be fitted across the 4" of the microfiche, meaning that the microfilm is scanned at 500 dpi (2,000 pixels divided over 4 inches).

The effective resolution has been calculated for each project image and reported in *Chart of Effective Resolution* in the **Appendix**.

Achieving successful resolution depends on a combination of the size of the map, the size and color of smallest type, and the complexity and range of the color coding. Lower resolution (normally less expensive to attain) will suffice for smaller and simpler maps. This is clear in comparing the map of Herkimer with the map of Mooers Quadrangle.

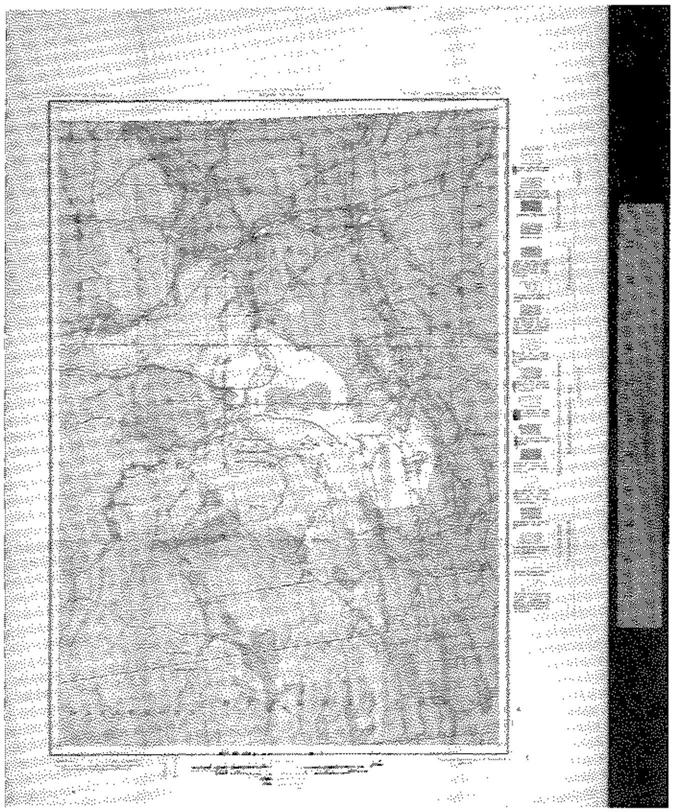
The 13.5"x9" map of Herkimer is very simple. It has five codes in one color (light brown) that differ by pattern and intensity and are easily distinguished from each other. The smallest type is 1.2 mm; all type is printed in black and quite easy to read in the original.



Map of Herkimer, Dimensions 13.5" x 9", scanned version legible at 130 dpi, 2048 x 3072 pixels

In set 3B (made from second generation microfiche, without color bars; see *Description of Vendor Products* in the **Appendix**), the smallest type on the Herkimer map is just barely legible on screen at a resolution of 1,024 x 1,536 pixels, and the five patterns are distinct even though the fine, parallel horizontal lines of the first code cannot be distinguished and the area appears to be a solid color. At this resolution the map is being viewed at approximately 66 dpi. At 2,048 x 3,072 vertical x horizontal pixels all the fine details of the map are fully visible. The effective resolution is 132 dpi, and the display is actually larger than the original map. In set 5B at 2,346 x 1,465 pixels the map is being viewed at 174 dpi and looks extremely good.

The Mooers Quadrangle map, in contrast, is larger and more complex. It is 17.5"x24", a typical full-color geologic map with fine detail and multiple color coding. It uses fourteen different codes in varying patterns and intensities of orange, green, pink, beige, and gray. Contour lines with elevations printed in 1 mm beige type are superimposed on top of the color coding. They are difficult to read on the original map without a magnifying glass.



Map of Mooers Quadrangle, dimensions 17.5" x 24", scanned version legible only at 175 dpi, 4096 x 6144 pixels

In set 3B, at a resolution of 2,048 x 3,072 pixels the smallest type on the Mooers map is not legible on screen, and it is just barely possible to read the labels in the key to the patterns. There is some difficulty distinguishing among the fourteen code patterns. The map is being viewed at approximately 87 dpi. At 4,096 x 6,144 pixels the contour elevations can be read and all codes can be fully distinguished. At 4,096 x 6,144 pixels it is being viewed at approximately 174 dpi effective resolution.

Due to the difference in size, the Mooers map must be viewed at double the resolution to reach about the same dpi as the Herkimer map. It appears that something in the range of 130-170 dpi on the paper original may be sufficient for both maps to be usable, although not for the finest details to be legible. The formula cannot be based simply on the size of the map and the size of the smallest type, however. The fact that the Mooers map is visually also much more complex, and not merely that it has smaller type than the Herkimer map, means that Mooers requires higher resolution to make the finer details fully legible. This is especially true where the lack of contrast between the elevations and the background plays a role. If Mooers were as simple as Herkimer, a somewhat lower resolution might suffice.

While some urge capturing all images at the highest technically possible resolution, the results of this project indicate that full information can be captured at lower resolutions depending on the size of the smallest type and on the pixel depth. In this case where we are using 24-bit color and capturing a 1 mm high letter, we need 200 dpi for full legibility. Someone scanning 1 mm high black and white text in binary mode would need 600 dpi for equivalent legibility. In either case, once the full information is captured at the highest resolution required for full capture, it should not be necessary to rescan in future, since no information would be added by scanning at the higher resolution.

These findings are in line with the proposals on required resolution set forth by Cornell for gray scale and color materials.[7] Cornell has proposed formulae for binary (2-bit) and grayscale (8-bit) scanning to derive benchmarks for recommended dpi capable of giving resolution equivalent to that of microfilm at quality index level 8 (the level required by the RLG). Following the logic of their formulae, a formula for color scanning (24-bit) suggests 200 for quality index level 8.

Based on the results of this project, it appears that a resolution of approximately 200 dpi in scanning the original captures all distinctions made on the original in full clarity. For purely access purposes on the other hand, display at lower resolution in the 130–150 dpi range leaves the larger type legible and allows for browsing.

It should be noted that the file of the 4,096 x 6,144 pixel version (i.e., 200 dpi) of the original Mooers map is over 18 megabytes in size, so large that it is not easily accessed and manipulated by many computers. At 17.5"x24", Mooers is not an enormously large map. The clear conclusion is that even larger maps with equally small print would require extremely large files in order to capture the print legibly. Tiling is one potential solution to this problem that was proposed by one vendor but not completed. It bears further investigation.

Color

All scanning was done at 24-bit color, where every pixel includes 8-bits of information each for red, green, and blue. The combination of the 24-bits of information potentially produces 16 million different color combinations, allowing for great richness of hues and extreme accuracy of reproduction if the scanning equipment is properly calibrated.

In order to determine whether colors would shift as a result of scanning, the vendors were requested to include standard color and 18-step gray-scale charts in scans of the originals. The charts were also included when the transparencies were made. Color and grayscale bars were not incorporated in the microfiche of the 1991-1992 project, unfortunately.

As noted above, color had shifted in all screen versions and all printouts. Shifts might be expected in the case of the transparencies and microfiche, since change is introduced by every aspect of the photographic process: the specific film stock, lighting, developing, printing. Yet shifts also occurred in the images printed from the originals. Even in printouts from the same medium from the same vendor, differences among the color bars could be distinguished. It is at least theoretically possible, however, to calibrate scanners, monitors, and printers by defining each of the 16 million colors relative to a standard, as long as standard color charts are included in the process. As long as standard color charts are present, use of color manipulation software can assist in bringing the colors closer to the original. This should not be considered enhancement but, rather, a way to reduce the amount of interference in the capture of the original color. It should be noted that such manipulation takes time and therefore increases costs.

Given the nature of these maps, the information conveyed by the color codes is adequately captured as long as the codes remain distinct, even though the actual hues may be altered. When considering the preservation of an accurate copy of the original, however, more work is needed to ensure that the color has been captured precisely; developing methods for accurate color capture, display, and printing is obviously a priority.

One complication in connection with the color charts should be noted. Because the placement of the color and grayscale bars was not specified, vendors used different arrangements. Some were more careful to maximize the use of space in the placement of the bars and orientation of the maps, while others left significant empty space around the maps and bars. Different scans of an individual map are therefore not fully comparable because the map may take up more or less of the image field at the same resolution depending on how much space is filled by color bars and margins. A map that fills the field is at a higher resolution than the same map when it fills only three-quarters of the field. Although cropping of the blank margins is possible, it requires manual intervention.

In sum, color and grayscale bars are important for calibrating color and minimizing shifts, but they take up space and pixels. The smallest available bars should be used, and they should be placed so as to minimize waste of space. In like manner, when the map is scanned it should be oriented to fill the image field as fully as possible. In presenting the images to users it would be possible to crop the color charts to save space, but the color bars should certainly be kept intact in the master file and made available to users who want to evaluate and/or manipulate the color they receive on screen.

Paper Printouts

The quality of the paper printouts from Iris machines provided by both Luna and Micro-Color is quite adequate. Fine details down to the smallest 1 mm numbers can be read on all of the printouts, and although colors shifted, all coding remained distinct.

The Iris printouts created for this project are strictly for use purposes and do not meet preservation criteria. The two types of paper used, a light-weight matte paper and a heavier glossy paper, are not acid-free. The pH averages 5.3–5.8. The inks are water soluble, and there is no information available about how long the inks last before fading. A comment from Micro-Color indicates that the manufacturer may be working on dyes with greater stability but no details were available. It would also be interesting to see whether the printers can accept acid-free papers.

The 11"x11" prints made by the Kodak Photo Printer XL7720 on Kodak electronic print paper are surprisingly legible, even though in the case of the two large maps the prints are less than a quarter of the size of the originals. Under a magnifying glass the finer details remain readable. Again, the paper is not preservation quality (pH is 5.5) and the inks are not stable; however, they are a very viable use medium.

Finally, a number of printouts were produced on various desktop printers. The Apple Color Style Writer 2400 printouts were reasonably legible, although limited to 8.5"x11" sections of the maps. Other lower-end printers were able to provide enough detail to suffice for browsing or shorthand copies.

User Evaluation

Evaluation of the vendors' products was carried out on several levels. At Columbia the project team, other librarians and computer experts, and faculty of the Department of Geological Sciences evaluated both the on-line images and the printouts. The on-line images and the paper printouts were shown away from Columbia several times and audience reactions recorded. Demonstrations were presented to attendees of the Geologic Society of America (GSA) conference in October, at which the Geoscience Information Society (GIS) also meets.[8]

Files Accessed Over the Internet

One project goal was to make the digital images of the maps available on the Internet to see if this is a viable way to give scholars access to the maps. A narrative description of the project accompanies the map images, and includes this report. The files are mounted on Columbia's World Wide Web server, where they can be accessed by anyone with Mosaic or Netscape capability at the URL:

http://www.columbia.edu/imaging/html/largemaps/

The existence of the images was announced on a variety of listservs read by librarians, geologists, and others interested in digital imaging. The people who responded were in a variety of disciplines interested in images, including, for instance, an experienced network user in medical informatics.

As discussed, the high resolution scanning resulted in files of 18–24 megabytes or more. Such large files cannot be accessed and manipulated except with high-end equipment. Therefore, in order to make the files usable on-line, Columbia has provided versions at lower resolutions accessible to most users. The on-line display states the files size for each image so that users can avoid trying to call up images that are too large for their equipment. For some images two versions are available, one with 256 colors in lossless GIF format and the other with full 24-bit color but using lossy JPEG quality 50 compression. Examples of versions that have been made available on-line are:

vertical x horizontal pixels	approximate file size
512 x 768 GIF	200-250 KB
1024 x 1536 GIF	1.5 MB
1024 x 1536 JPEG	77 KB
2048 x 3072 PhotoCd	5-8 MB

While many people found retrieving the maps easy, there were unfortunately—but not unexpectedly—some problems with remote access. One of the Columbia WWW servers was down most of the week that the URL was announced publicly, and this caused many people to be unable to access it or to have to try several times before succeeding. Even after this problem was solved, the usual facts of Internet life remain as barriers to consistent, easy use of images.

Access through the WWW takes time. Even on the Columbia campus when traffic is heavy it can take several minutes to load the large file of a high resolution image, and accessing it remotely can take hours longer, especially when the Internet is busy. The smaller, lower resolution files move faster but cannot provide all the details. An obvious solution would be to determine the lowest resolution for a successful browsing tool and mount that as the file recommended for real time access, with high resolution versions available for downloading during low-use periods.

All of the evaluators were asked to determine which images (both scanned and printed out) were acceptable or not acceptable for identification and research purposes. The responses to the various images were overwhelmingly positive. Evaluators were generally enthusiastic about the quality of resolution, which made all but the largest and most complicated map (with some of the smallest print) fully legible on-line. There was agreement that even the lowest resolution image of the map of Herkimer at 512 x 768 was a total success for browsing purposes when viewed on a variety of monitors.

Viewers observed that the more complex and larger maps could not even be browsed at low resolution. At the same time, the higher resolution images of the large maps could not be viewed as a whole at the screen. Inability to view the whole surface can sometimes be disorienting when trying to read a map, and studies have indicated that peripheral vision is very important in map use.[9] Several people suggested sectioning the maps as an intermediate option; others did not want to give up the ability to move around the whole map within one file even if only part of it was visible at one time. Viewers concurred that although we might not be able to view the whole large images today, it was just a matter of time before desktop computing capacity caught up.

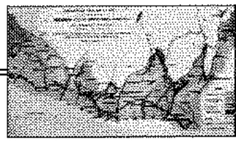
Other aspects of on-line use elicited less favorable comments. Viewers were concerned about access difficulties and the time and memory size needed to view the maps. While many in the geologic sciences are not interested in the quality of the color (as long as they can distinguish easily among the codes), a few commented on the color shifts as an issue that should be addressed. There was some discussion about scale. Most maps carry a scale indicating what distance is represented by an inch or centimeter, and some evaluators felt it would be helpful to add a note of reminder to viewers that the size of the map was not necessarily one to one at the screen or in a printout. Whether or not a scale was explicitly printed on the map, evaluators stated that it would be very important to include a ruler in the image so that the original size could be easily determined (since the color bars include a ruler, they thus can serve two purposes). A final issue was the degree of distortion that may have been introduced through scanning. Use of lenses in photography is known to distort the image's edges and investigation was urged into whether scanning directly from the original might eliminate the potential for distortion.

Printouts

Evaluators viewed the three sets of Iris prints as well as the 11"x11" photo prints and a sampling of 8.5"x11" printouts from a standard desktop color printer. They were generally quite happy with the Iris and the 11"x11" prints as use media. While the small prints were seen as less useful than the full-size printouts for serious research, a number of evaluators considered them a handy and inexpensive intermediate product that could fill needs for identification and browsing. Evaluators were also happy with the option for local desktop printing. Although desktops produce only 8.5"x11" sections of the maps and have limited color ranges, the printouts provide a very convenient sketch of the map for identification purposes for anyone who does not need the full-color, full-size version.

Evaluators were divided in their preference for matte versus glossy paper and agreed that in would be good to have both types. They noted that some of the Iris prints were not sized exactly to the originals and urged that this be corrected in future. Again, printing the map with a ruler to the side would help in verifying the scale used in the map. Several also noted the color shifts and hoped that this could be corrected in future as well.

Printout longevity was a concern. Many evaluators said that it would be highly desirable to be able to produce printouts on demand equivalent in permanence to preservation photocopies with which to replace out-of-print damaged and brittle originals. Several asked how soon Columbia would be able to make copies available on demand and indicated immediate interest in purchasing sets of the maps.

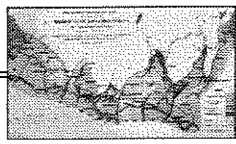


Bibliographic Control

As part of the project it was decided to experiment with bibliographic control for the digitized maps. While the Internet allows wide access to electronic resources, it lacks any coherent type of indexing to let people know what resources are available. Since the national databases serve this function well, it is important to add MARC records for digital files to RLIN/OCLC in order to alert researchers to their existence.

Guidelines and standards for creation of records for digital copies of paper materials are far from established, and Columbia is working actively to develop them with RLG, OCLC (through its Internet Cataloging Project), and the Intellectual Access Committee of the American Library Association's Preservation and Reformatting Section. Issues still to be decided include what data about images should be recorded and where, i.e., what portions of this information might be part of the bibliographic record, which should be recorded with the image elsewhere. Useful and necessary information includes: the type of film stock used for photographic versions that have been scanned, the reduction ratio, and the dimensions of those photographic versions. Further data concerns the method of scanning, any color correction or other manipulation, equipment use and software used, storage media, and so forth.

In the meantime, for this project a collection-level bibliographic record for the project as a whole has been created. The digitized images are described and subject access is provided. The record also refers people to the Web site for access to the files and the narrative report on the project. A 5xx note provides a narrative reference, and the URL is also entered into the 856 subfield *u* where eventually it will constitute a hyperlink. See *Bibliographic Record for Project* in the **Appendix**.



Costs

One area that needs much more attention is the basis for predicting costs. Because this was such a small project, several vendors donated their services, and those that charged for the work did not have a production-level basis on which to estimate realistic costs for services they do not normally provide. In any case, the pricing of digitization services and equipment is constantly in flux. As a result, it is not possible to draw conclusions about relative prices from this project. We can simply report the range of prices charged:

scanning the original maps	$75.00 each
scanning the microfiche	$25.00-32.00 each
scanning the transparencies	$28.00 each
Iris prints of 2 larger maps	$145.00-250.00 each
Iris prints of 3 smaller maps	$40.00-55.00 each
11"x11" prints	$8.00 each

There were also a variety of separate charges for writing files to the storage media (PhotoCD, tape, etc.) that ranged from $4.00-20.00 per file.

Clearly, there is a need for cost models for planning in a full production level digital project. Yale and Cornell's work on what it takes to ramp up will be very important in trying to get an idea of what digital imaging really costs.

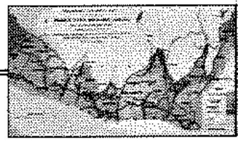
Working with Vendors

This project has made clear the need for explicit communication between the library and its vendors. The documentation provided by the vendor is a key part of a successful project. Otherwise, identification of files and determination of exactly what resolution, compression, etc., were used becomes extremely difficult. Some vendors voluntarily provide detailed documentation; in other cases the library will need to specify what information it wants from the vendor and in what form. The information contained in the file header is not sufficient. Questions of how to record data about images and intermediaries have begun to be discussed by libraries, and once a common set of guidelines is established it will be much easier to specify for vendors what information is needed and in what form.

Simply conveying to vendors what services and product quality are desired is not always easy. At present there is no convenient and accepted standard, which makes it necessary to spell out in detail what is wanted. Communication is hampered by a lack of shared vocabulary and clear, standardized language to describe what is needed. There is also a lack of understanding by librarians as well as some of the vendors as to what some technical terms mean, complicated by the rapid pace of change industry-wide.

Because quality control procedures for vendors and especially for the library are not yet established, it is not always clear how to guarantee that the end result from vendors is what was requested. It can also be difficult to negotiate rework with a vendor if the resulting product is unsatisfactory.

The solution is partly a matter of becoming educated in digital imaging literacy and partly a matter of finding vendors who are interested in such work. As was (and is) the case with microfilming, most vendors have no idea what preservation is and what the goals are. Some vendors are more concerned with selling the product they already produce rather than tailoring their operations to preservation needs. It is important to continue to identify vendors who are interested in learning about and willing to work on the development of preservation-quality services.



Organizational Issues

Digital projects are extremely complex to design and manage. They require expertise not normally found in a single library department or division and sometimes not found in the library at all. Such projects are beyond what a single curator or a single preservation department can afford to tackle alone. Issues such as decisions on institutional policies, commitment of technical and hardware support, provision of financial support, planning for the future, assurance that digital files will be maintained, refreshed, and migrated, must all be made on a system-wide basis.

Columbia uses a team approach to digital library issues. A standing committee has been established by the Columbia University Libraries to oversee the progress of the digital library, coordinate projects, and discuss issues in depth and come up with decisions. The committee is headed by the Deputy University Librarian and is composed of: the Director of Library Systems; Academic Information Systems (formerly Academic Computing) staff responsible for coordinating faculty and library imaging projects and mounting them on the Columbia Web server; the Director for Bibliographic Control; curators of collections in which projects are currently taking place or who have a high degree of experience with digital issues; and the Director for Preservation.

Each individual project has also a management team that brings together expertise on the subject matter and the needs of users, preservation, and the technical side of imaging. This particular project has a three-member team from Preservation, Geology, and academic Information Systems. We have found that the team approach works very well and recommend it to other institutions.

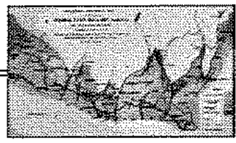
Next Steps

Phase II of the project, an attempt to create an integrated on-line version combining text and illustrations, is now underway. The project is assembling a full film copy of four volumes of the *Museum Bulletin*, using a combination of microfilm of the text and single-frame color microfiche of the illustrations; then scanning the microfilm of the text at 600 dpi black and white; and scanning the microfiche of the illustrations at approximately 200 dpi, 24-bit color. This will produce a series of digital files, one for each page and for each illustration. The project 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 and recreate the original volumes virtually on the Internet.

The result of the project will be the full preservation of four sample volumes. Long lasting microfilm/fiche will have been created for all the text and illustrations; this will be done in a truly usable digital version where the author's juxtaposition of words and illustrations is also preserved and available for users to read on-line or to create paper printouts.

A number of other areas have been identified as needing further research. It will be very important to determine the real costs of digital imaging, so that decisions on what endangered materials to convert and when can be made on a firm financial basis. Guidelines for adequate and high-level quality of resolution for various media and for various levels of use must be developed, along with schemes for institutional quality control evaluation to ensure that required levels of quality have been provided by vendors and in-house scanning services. Improving the accuracy of color capture and display is another obvious area for research if digital imaging is to serve disciplines such as art history, where faithful representation of color is essential. Finally, the potential for tiling as a way to deal with very large images needs to be investigated. How difficult is it to achieve exact registration when moving from one section to the next? Would scholars find this a useful method for working with large images? How would tiling affect the role of peripheral vision in the use of large images?

The use of digital imaging for preservation purposes offers exciting possibilities. While much further research awaits, it is our hope that this project on the digitization of oversize color images will help to answer some of the questions.



Appendix

Description of Vendor Products

The following describes the work performed by each vendor, the equipment used, the total number of vertical times horizontal pixels used for the image (not to be confused with dots per inch), compression method, and comments. All of the scanned images are available over the Internet. The set numbers in the descriptions here refer to the on-line images accessible on Columbia's World Wide Web server URL: http://www.columbia.edu/imaging/html/largemaps/

Boston Photo (set 1 in the on-line version)

Boston Photo worked only with the transparencies and did not produce paper printouts.

set 1a. Produced 35 mm slides from the transparencies and scanned them on a Kodak PCD 2000 scanner to Kodak Photo CD at a maximum resolution of 2,048 x 3,072 (vertical x horizontal pixels). Because of the loss of quality in creating a derivative slide, this set of images was not considered successful.

set 1b. Scanned the transparencies directly on a Kodak PCD 4045 scanner to Kodak Pro Photo CD at a maximum resolution of 4,096 x 6,144 pixels and 24-bit color.

The compression method used in both cases was the proprietary Kodak Photo CD format.

Micro-Color (set 2 in the on-line version)

set 2a. Scanned directly from paper originals using the Leaf Digital Studio camera at a resolution of 2,048 x 2,048 pixels and 24-bit color, wrote the files to Kodak PhotoCD, and also stored them as uncompressed TIFF files.

set 2b. Created new first generation microfiche and scanned them (not the 1991-1992 fiche) using an Optronics drum scanner at resolution of 2,000 dpi (ca. 4,000 x 5,000 pixels for the two larger maps and 2,500 x 3,500 pixels for the three smaller maps) and 24-bit color. The files were written to Kodak PhotoCD and also stored as uncompressed TIFF files.

The compression method used in both cases was the proprietary Kodak Photo CD format.

2c. Printed out from both using Iris Smartjet 4012 for the three smaller maps and the Iris 3024 printer for the two larger maps. All were printed on coated paper.

Luna (set 3 in the on-line version)

set 3a. Scanned from transparencies to Kodak Pro PhotoCD using resolution of 4,096 x 6,144 pixels and 24-bit color and Kodak PhotoCD compression. Also produced uncompressed RGB TIFF files. Scanned the Q60 standard color chart and used it to color-correct the images based on colorbars in the transparencies.

set 3b. Scanned from second generation 1991-1992 microfiche to Kodak Pro PhotoCD using resolution of 4,096 x 6,144 pixels and 24-bit color and Kodak PhotoCD compression. Also produced uncompressed RGB TIFF files. Attempted color-correction of the images made from the microfiche by correlating to the color of the images made from the transparencies where color bars were included.

Luna used the Kodak 4200 Professional Imaging System, consisting of:
- Kodak 4045 film scanner
- Sun Sparc2 Workstation
- Kodak CD ROM Writer 200
- Kodak Professional Image Capture Software Version 4.2

Luna also provided a diskette listing image numbers, conversion types and date, resolution and bit-depth, rotation, nature of the item being scanned, and related information.

3c. Printed out from both sets of images onto matte paper using an Iris 3047 printer.

3d. Printed out 11"x11" prints from both sets of images using the Kodak Photo Printer XL7720 onto Kodak electronic print paper.

Stokes (set 4 in the on-line version)

Stokes intended to use a digital camera to scan the original maps in sections at high resolution and then concatenate the sections to produce one large image (tiling).This plan was based on their expectation of acquiring appropriate hardware and software in time for the project. By that time they were not yet equipped to concatenate the sections and handle the huge files that would have resulted. Instead, Stokes scanned each original map in three ways.

First, the entire map (up to 14.63"x19.63") was scanned at 2,984 x 2,312 pixels using a Kontron digital camera. Second, a 9"x12" section was scanned at the same resolution, and finally a 6"x12" section at the same resolution. As the section grows smaller, the effective resolution grows higher;[10] for instance, the resolution of the entire Catskill map is 204 dpi, the 9"x12" section is 270 dpi, and the 6"x8" section is 373 dpi.

JPEG compression was used. It was written at a quality factor of -5 (compared to the highest quality at -7).

The second generation (1991-1992) microfiche and transparencies were sampled to produce scans of the maps and the sections that matched the scans from the original in dpi. The 6"x8" section of the original is captured on a 1"x1.25" area of the fiche (and also the transparency). That area was scanned at 2,387 pixels, which translated into a resolution of 373 dpi on the original. In other words, the scan of the 6"x8" section of the original, the 1"x1.25" area of the fiche, and the 1"x1.25" area of the transparency when viewed side by side are all the same size and the same resolution.

The sectioning does allow for much higher resolution views of the maps. Unfortunately, a number of the images are somewhat out of focus along the right side. This appears to be correctable through rescanning.

Visual Information (set 5 in the on-line version)

Scanned the original maps and the microfiche. Both Tangent and Sharp flatbed scanners were used. Smaller maps were scanned at 300 dpi and larger ones both at 300 and 400 dpi. All scanning was done at 24-bit color.

Wrote the images of the three smaller maps to a CD-ROM containing their proprietary ImageBASETM software and wrote the images of the two larger maps to 8mm tape. The files were also provided in uncompressed TIFF format.

Printed out the Buffalo map from a reduced resolution image using a Hewlett-Packard color printer, onto matte paper that was then laminated.

Chart of Effective Resolution (Approximate)

The dimensions of the map plus color bars and margins are given under each map name. Note that dimensions change from set to set because vendors placed color bars in different places and left different sizes of margins. Effective resolution has been calculated using these dimensions.

Most of the images in sets 4A-4C consist of sections rather than the entire maps; only the first group in set 4A contains images of the entire maps and is included here.

Set 1B — transparencies shot with color bars

	display size in horizontal x vertical pixels			
	512x748	**1024x1536**	**2048x3072**	**4096x6144**
Perce (23.25"x12.25")	22 dpi	44 dpi	88 dpi	176 dpi
Catskill (16.5"x18.5")	31 dpi	62 dpi	124 dpi	248 dpi
Herkimer (20.75"x12.5")	25 dpi	49 dpi	99 dpi	197 dpi
Mooers (25"x27.5")	20 dpi	41 dpi	82 dpi	164 dpi
Buffalo (24"x25")	21 dpi	43 dpi	85 dpi	171 dpi

Set 2A — original maps shot with color bars

	display size in horizontal x vertical pixels			
	512x748	**1024x1536**	**2048x3072**	**4096x6144**
Perce (16"x14.75")	32 dpi	missing	missing	missing not provided
Catskill (17"x15")	30 dpi	60 dpi	120 dpi	missing not provided
Herkimer (17"x15")	30 dpi	60 dpi	120 dpi	missing not provided
Mooers (25.5"x24")	20 dpi	40 dpi	80 dpi	missing not provided
Buffalo (25.75"x21.5")	20 dpi	40 dpi	80 dpi	missing not provided

Set 2B — first generation microfiche shot without color bars

	display size in horizontal x vertical pixels			
	512x748	**1024x1536**	**2048x3072**	**4096x6144**
Perce (14"x8.75")	37 dpi	73 dpi	146 dpi	293 dpi
Catskill (10.25"x15")	50 dpi	100 dpi	200 dpi	400 dpi
Herkimer (14.5"x9")	35 dpi	71 dpi	141 dpi	282 dpi
Mooers	not available			
Buffalo (20"x25")	26 dpi	51 dpi	102 dpi	205 dpi

Set 3A — transparencies shot with color bars

	display size in horizontal x vertical pixels			
	512x748	**1024x1536**	**2048x3072**	**4096x6144**
Perce (23.25"x12.25")	22 dpi	44 dpi	88 dpi	176 dpi
Catskill (16.5"x18.25")	31 dpi	62 dpi	124 dpi	248 dpi
Herkimer (20.75"x12.5")	25 dpi	49 dpi	99 dpi	197 dpi
Mooers (25"x27.5")	20 dpi	41 dpi	82 dpi	64 dpi
Buffalo (23"x25")	22 dpi	45 dpi	89 dpi	178 dpi

Set 3B — second generation microfiche shot without color bars

	display size in horizontal x vertical pixels			
	512x748	**1024x1536**	**2048x3072**	**4096x6144**
Perce (14.5"x8.75")	35 dpi	71 dpi	141 dpi	282 dpi
Catskill (11.50"x15")	45 dpi	89 dpi	178 dpi	356 dpi
Herkimer (15.5"x9")	33 dpi	66 dpi	132 dpi	264 dpi
Mooers (23.5"x24")	22 dpi	44 dpi	87 dpi	174 dpi
Buffalo (21.5"x21.5")	33 dpi	66 dpi	132 dpi	264 dpi

Set 4A — original maps shot with color bars

	display size in horizontal x vertical pixels	
	1492x1156	**2984x2312**
Perce (15"x12.25")	99 dpi	199 dpi
Catskill (12.5"x15")	119 dpi	239 dpi
Herkimer (14.5"x11.25")	103 dpi	206 dpi
Mooers (margins trimmed) (15.5"x24")	96 dpi	192 dpi
Buffalo (margins trimmed) (16"x21.5")	93 dpi	187 dpi

Set 5A — original maps shot with color bars

	display size in horizontal x vertical pixels
	1640x1320
Perce (16.5"x12.25")	99 dpi
Catskill (13.75"x18.5")	119 dpi
Herkimer (18.5"x12.5")	87 dpi
Mooers (21.5"x27.5")	76 dpi
Buffalo (23"x25.5")	71 dpi

Set 5B — second generation microfiche shot without color bars

display size in horizontal x vertical pixels (each map's display differs)

	1202x850	**2404x1700**
Perce (12.5"x8.75")	96 dpi	192 dpi
	820x660	**1622x2745**
Catskill (9"x15")	91 dpi	180 dpi
	820x660	**2346x1465**
Herkimer (13.5"x9")	61 dpi	174 dpi
Mooers (21.5"x27.5")	not available	
Buffalo (23"x25.5")	not available	

Bibliographic Record for Project

collective record for the digitized images

```
LTCU DONE                                                    AKD8206
                                              NOTIS CATALOGING          2W99
CU- AKD8206 FMT U RT b BL m DT 01/12/95 R/DT 01/12/95 STAT nn E/L 5 DCF a D/S S
SRC d PLACE nyu LANG eng MOD    REPRO    D/CODE i DT/1 1994 DT/2 1995

040:   : a NNC c NNC
043:   : a n-us-ny
245:10: a Geological maps, New York State h [computer file] f 1905-1906.

300/1:   : a 15 image files : b GIF
520/1:   : a Digital images of five geological maps extracted from New York
State Museum Bulletin ; 80, 84, 95, 99. b Images scanned by a variety of
vendors using different methodologies, including scanning from originals,
transparencies, 35 mm slides, color microfiche.  Images available at varying
resolutions, with file sizes of 0.5 MB to 6 MB.
590/2:   : a Digital images, including detailed description, available via
Internet: URL=http://www.cc.columbia.edu/imaging/html/largemaps/oversized.html
650/1: 0: a Geology z New York (State)
730/1:0 : a Bulletin (New York State Museum)
```

collective record for the paper maps, treating the digital images as reproductions

```
LTCU DONE                                                    AJW7029
                                              NOTIS CATALOGING          2W99
CU- AJW7029 FMT U RT b BL m DT 10/05/94 R/DT 01/12/95 STAT nn E/L 5 DCF a D/S S
SRC d PLACE nyu LANG eng MOD    REPRO    D/CODE i DT/1 1905 DT/2 1906

040:   : a NNC c NNC
043:   : a n-us-ny
245:10: a Geological maps, New York State, f 1905-1906.
300/1:   : a 5 maps : b col. c 83 x 57 cm. or smaller
500/1:   : a Extracted from: New York State Museum bulletin ; 80, 84, 95, 99
583/2:   : a Reformatted, c 1994 i digital image
520/3:   : a Geological maps extracted from issues of New York State Museum
bulletin as part of a project to test scanning methodologies for large-scale
color images. b Maps range in size from 8.75" x 12" to 17.5" x 24" and include
a variety of color tones and fine detail.
530/4:   : a Available, with detailed description, as digital images via
Internet: b
URL=http://www.cc.columbia.edu/imaging/html/largemaps/oversized.html
650/1: 0: a Geology z New York (State)
730/1:0 : a Bulletin (New York State Museum)
```

Endnotes

[1] *Scholarly Resources in Art History: Issues in Preservation, Report of the Seminar, Spring Hill, Wayzata, Minnesota, September 29-October 1, 1988* (Washington, DC: Commission on Preservation and Access, 1989).

[2] Joint Task Force on Text and Image, *Preserving the Illustrated Text* (Washington, DC: Commission on Preservation and Access, 1992).

[3] Preservation Department, "Report on the Condition of the Collections" (Chicago: University of Chicago Library, 1989).

[4] The Image Permanence Institute of the Rochester Institute of Technology is currently more than half way through a two-year project sponsored by the State of New York, "Isoperms for Color Photography," which will provide the data to establish the life expectancy of color photographic film in optimum storage conditions. Cibachrome/Ilfachrome microfiche produced and stored according to national standards is already known to have at least a 100–200 year life expectancy.

[5] Single-frame color microfiche use the entire surface of the microfiche for a single image, thus allowing for a significantly lower reduction ratio than microfilm, so that details can be read without sectioning the map.

[6] These vendors were: Boston Photo Imaging, Boston, MA; Luna Imaging, Inc., Venice, CA; Micro-Color International Inc., Midland Park, NJ; Stokes Imaging Services, Austin, TX; Visual Information Inc., Denver, CO.

[7] Anne Kenney and Stephen Chapman, *Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality* (Washington, DC: Commission on Preservation and Access, 1995), p. 10.

[8] For reports on these demonstrations see: "Pilot Project Demonstrates CD-ROM, Fiche Products," *The Commission on Preservation and Access Newsletter*, February 1993, pp. 2-3; and "Comments Sought on Oversized Map Project," *The Commission on Preservation and Access Newsletter*, February 1995, pp. 3-4.

[9] See, for instance, Richard Phillips, "Making Maps Easy to Read," (URL http://acorn.educ.nottingham.ac.uk//ShellCent/maps/) on a project conducted at University College London at the Royal College of Art and at the University of Nottingham.

[10] See *Project Results: Effective Resolution* in this report for a discussion on this topic.