



[CONTENTS](#)

[SEARCH](#)

[HOME](#)

August 15, 1998, Volume 2, Number 4, ISSN 1093-5371

Table of Contents

- **Feature Article**
 - [Creating and Distributing High Resolution Cartographic Images](#) by David Yehling Allen
- **Technical Feature**
 - [Imaging Production Systems at Corbis Corporation](#) by Sabine Süsstrunk
- [Highlighted Web Site](#)
- [Calendar of Events](#)
- [Announcements](#)
- [FAQs](#)
- [RLG News](#)
- [Hotlinks Included in This Issue](#)

Feature Article

Creating and Distributing High Resolution Cartographic Images

David Yehling Allen

Map Collection, Melville Library, SUNY Stony Brook

dyallen@notes.cc.sunysb.edu

Introduction

Maps and other cartographic materials are popular items for digitization and display on the Internet. Antique or historical maps, contemporary maps, aerial photographs, and satellite images can all be found on the World Wide Web. This article will focus on the reformatting of paper maps as computer-readable

images, both for purposes of preservation and for access through CDs or on the Web.

Maps digitized from originals on paper almost always need to be in raster format, in which the image is formed by a grid of dots or pixels. Many of the maps found on the Internet are vector maps, which are created from mathematically defined lines and curves. Vector maps are usually produced by Geographic Information Systems (GIS) software, and are typically used in interactive programs, such as the Census Bureau's [TIGER Map Service](#). Vector maps are more versatile than raster images. Vector maps can be easily rescaled; they can be made to display information in multiple layers; and they can be used to create thematic maps derived from geocoded information contained in databases. But the vector format is not suitable for digitizing paper maps, especially historical maps. The wide variety of fonts, images, and lines on antique maps are virtually impossible to translate into vector images.

The quality of raster images on the Internet varies greatly. Many can best be described as "evaluation quality" images - they give a good idea of the overall appearance and contents of a map, but do not preserve enough detail for many research purposes. "Research quality" images preserve all of the significant details on maps (such as street names and contour lines). "Preservation quality" images are scanned to the same or higher resolution as research quality images, but are more carefully controlled for image quality and consistency. There are no widely recognized standards for "research quality" or "preservation quality" images, but *de facto* standards are beginning to emerge, at least for research quality images, and will be discussed below.

For those who want to sample the wide variety of cartographic sites on the Web, the following are good starting points. The largest collection of small scale reference maps is probably at the [Perry-Castenada Map Library at the University of Texas at Austin](#). Typical collections of "evaluation quality" images of historical maps can be found at the [Hargrett Library at the University of Georgia](#) (maps of Georgia and the Southeast); the [University of Arizona](#) (maps of the Southwest); [SUNY Stony Brook](#) (maps of New York); and the [Osher Map Library at the University of Southern Maine](#) (maps of New England). Those who want a comprehensive listing of digitized maps on the Internet can find one at [Odden's Bookmarks](#). Few of the maps at these sites are digitized at a high enough resolution to qualify as research quality images.

Creating High Resolution Images of Maps

There is by now a good deal of consensus as to what does constitute a research quality image. An important study conducted at Columbia University, the [Oversized Color Image Project](#), concluded that, in the case of most maps, all of the essential details could be captured at 200 dpi in 24-bit color. The [Library of Congress](#) has adopted a 300 dpi 24-bit standard for its map digitization projects, and this is likely to become the *de facto* standard for similar projects. The [National Archives](#) has adopted essentially the same standard, although concerns about file size have led that institution to scan maps larger than 11 x 17" at 200 dpi (1). Digitization at 300 dpi with 24-bit capture leaves some room for error, and even makes it possible in many cases to discern such details as paper texture. Although this standard does not satisfy all conceivable needs, very few researchers require more detail. It should be noted, though, that black and white maps do not necessarily have to be scanned in color. Gray scale and bitonal scanning are options for uncolored maps. As will be seen below, the National Oceanic and Atmospheric Administration (NOAA) is

doing bitonal scanning of black and white Coast Survey charts. When bitonal scanning is used, materials are scanned at a higher resolution (between 600-800 dpi) to gain image quality comparable to 300 dpi 24-bit color scanning, but the size of the bitonal files is still much smaller than the color.

It is actually quite easy to produce 300 dpi 24-bit color images using standard off-the-shelf equipment. At the [University of Connecticut](#), the map librarian has been scanning maps in sections on an ordinary commercial scanner at 300 dpi, and then tiling the images or knitting them together using Adobe Photoshop. The resulting images are quite presentable, and researchers viewing them over the Web can indeed see all of the significant details.

Still, such images do not satisfy the exacting requirements of some granting agencies, and cannot be depended upon to *consistently* meet the needs of researchers. To consistently deliver acceptable quality, a number of other criteria need to be met. These include the use of precision equipment, careful control over production processes, systematic evaluation of image quality, and adequate metadata. The technical issues involved in producing high quality images have been described elsewhere (2). Although the process is demanding and expensive, a number of institutions are producing digital images of maps that meet these criteria. These images are created using oversized scanners or expensive digital cameras, which can now produce images up to 10,000 by 10,000 pixels (3). Such devices can produce high quality digital images at 300 dpi in 24-bit color from maps up to approximately 30" x 40". Larger or irregular maps still have to be dealt with by tiling images, or by knitting them together with a program such as Photoshop.

It is debatable whether even the high-quality research images described above constitute "preservation quality" images. There is, of course, an ongoing discussion about whether digitization is a viable means of preserving anything. Much of this debate hinges on such issues as the durability of digital documents, and goes far beyond the scope of this article. More directly relevant are the issues of color appearance and image resolution. Complete color matching is not of much concern to typical map users, although colors need to be rendered accurately enough to distinguish between the different subjects on thematic maps, and to correlate map legends to content regions. Still, if a preservation quality image is to constitute a permanent record of a map, one might well refuse to settle for anything less than full color matching. Even more thorny is the question of how much resolution is needed for preservation purposes. Although 300 dpi is adequate for perhaps 99% of all current research, what about the remaining one percent? And what about possible future research projects that we cannot even foresee? Questions like these make it almost impossible to visualize a standard everyone will accept.

As a partial response to these questions, several government agencies scan maps at higher resolutions than they use for access purposes. The U.S. Geological Survey, for example, has been scanning its topographic maps at 500 dpi, and then resampling the images at 250 dpi for distribution. These are the Digital Raster Graphic (DRG) data files that the USGS has been sending to depository libraries on CDs. Similarly, both the National Archives and NOAA (in projects discussed elsewhere in this article) distribute scaled down versions of their images over the Internet. These efforts, however, have as yet led to no consensus as to whether it is really necessary to scan master images at a higher resolution than 300 dpi in 24-bit color, and no agreed-upon minimum resolution for such archival images appears to be emerging.

The most important practical problems confronting the producers of digital images of maps arise from the huge size of raster image files. A fairly ordinary sized map, such as a topographic map measuring 22"x 27", digitized at 300 dpi in 24 bit color, will produce an uncompressed file 160 Mb in size. It is not unusual for digitized maps to occupy files much greater than 200 Mb, and some individuals are producing files of more than 500 Mb for a single map. Images of this size can be compressed with little or no loss and distributed on CDs. Thus the Digital Raster Graphic maps mentioned above are being distributed as "pack bits compressed" TIFF files, which occupy about 12 Mb each. These large files can be compressed even further with little perceptible loss of information using emerging compression techniques. The advent of DVDs, which can hold several times as much data as CDs, should further facilitate the distribution of high quality digital images of maps (4).

Distributing High Resolution Cartographic Images on the Internet

One of the biggest challenges confronting the producers of high quality digital images is their distribution on the Internet. These gigantic files create multiple problems. They make heavy demands on server space. They take a long time to transmit, especially to those who do not have access to high-speed networks. And they make excessive demands on computers with little memory, limited hard drives, or inadequate displays. Once one of these large files is downloaded, only a small portion of the map can be displayed on normal monitors at any given time, and some users find it difficult to use maps without having the entire map in front of them.

In spite of these obstacles, great ingenuity has been shown in developing ways to transmit maps in digital form over the Internet. The remainder of this article will be devoted to exploring several sites that illustrate various approaches to this problem.

Perhaps the simplest approach is illustrated by a site managed by the National Oceanic and Atmospheric Administration (NOAA). NOAA has recently made available on the Web a spectacular [Historical Map and Chart Collection](#). As noted earlier, maps in this collection were originally digitized for archival purposes at 800 dpi. Because many of the original charts were black and white, much of this scanning was bitonal, with 24-bit color being reserved for colored maps. The file size of the 800 dpi bitonal images averages a relatively small 5-10 Mb; similarly sized color maps result in images that are 24 times larger. For distribution purposes, the user can view "evaluation quality" 100 dpi GIF images using a standard Web browser. Those who need higher quality maps can download 300 dpi TIFF images. The file size of most of the resampled bitonal images is under 1 Mb; the size of the color images is greatly reduced by resampling them to 8-bit color. This is a distinctly "low tech" solution, but it does provide reasonably good access to an outstanding collection (5).

Another "low tech" approach to distributing large map images is "tiling." A good example can be found at the [University of Connecticut](#) site, where the 1834 Tanner map of the United States has been divided into 60 sections or "tiles." Each section is approximately 500 Kb, and can be displayed by clicking on the corresponding tile on a low resolution index map. This approach suffers from two drawbacks. Creating such a tiled image is quite labor intensive, and tiled images are at best cumbersome for the user who needs to simultaneously examine information displayed on two or more adjacent tiles.

Tiling combined with other techniques can overcome some of the limitations of tiling alone. [The Atlantic Neptune Online](#) is well worth visiting for its creative use of tiling in combination with three different zoom levels.

Another innovative approach to distributing map images on the Internet has been taken by the National Library of Iceland with its [Icelandic Map Project](#). This ambitious project involves the creation of a comprehensive online library of historical maps of Iceland. Maps were scanned from 6 x 7 cm positive color slides using a Kontron film scanner to produce 3,500 x 3,500 pixel images in 24-bit color. The master images, which average about 35 Mb in size, are stored on CDs as TIFF files with LZW compression. Derivative images in PNM (Portable Anymap) format are stored on a server for access on the Web. These derivatives are accompanied by JPEG images in three resolutions for quick display on different screen sizes (640 x 480 pixels, 800 x 600 pixels, and 1,024 x 768 pixels). The resolution of the JPEG images can be doubled by clicking on them (this highest resolution is derived "on the fly" from the PNM derivative on the server). The images are accompanied by two thumbnails. One of the thumbnails is used for display with text. The second is displayed with the enlarged map images. Displaying the thumbnail together with an enlarged section of a map is a useful innovation, which makes it considerably easier for the user to comprehend the relationship between a particular portion of a map and the map as a whole. Unfortunately, even the highest resolution is not adequate to capture all of the significant details on large maps. The maximum resolution is constrained by the 3,500 x 3,500 pixel file size, and by the loss of resolution involved in converting film images to digital files (6).

The most interesting and promising solutions to distributing maps on the Internet involve innovative software programs. In my opinion, the best results so far have been obtained using the Multi-resolution Seamless Image Database (MrSID) software developed by a company called [LizardTech](#). This software combines high quality wavelet compression with a technology that allows users to zoom in on small portions of a map (7). MrSID is quite flexible, enabling users to examine a map at a variety of levels of magnification without the restrictions imposed by tiling. Because MrSID displays only as much of a map as will fit on the screen, only a small part of the image files needs to be downloaded at a time. The best known application of MrSID is the [Panoramic Map Project at the Library of Congress](#). It has also been used at a number of other sites, including: [The San Francisco Bay Area Orthophotoquad Project](#) of the United States Geological Survey; the [Illinois Aerial Photograph Digital Archive](#) at the University of Illinois at Urbana-Champaign; and a large historic map collection from the [Virginia Board of Public Works](#) at the Library of Virginia.

In spite of its impressive performance, MrSID has some drawbacks. It requires a browser "plug in," and downloading times are still significant. For those with powerful computers it is much easier to scroll around a single high-resolution image of an entire map than to examine it a bit at a time over the Internet using MrSID. LizardTech has addressed this latter problem by making available a free viewer, which works with downloaded images in MrSID format, and also allows one to convert these images into TIFF and other formats for editing. But not all institutions that make their images available in MrSID format permit the downloading of whole files, and this procedure may be too cumbersome for the casual user.

Software in this area is rapidly developing, and there are products on the market that might be used quite effectively with cartographic materials. An example is FlashPix, which was discussed in the April 15, 1998 issue of *RLG DigiNews* (8). FlashPix has zooming capabilities similar to MrSID (although it works with tiled images), but it can be used with a wide variety of file formats. To the best of my knowledge, FlashPix has not been used in any major cartographic imaging projects, but its capabilities are demonstrated by a test application using images supplied by the U.S. Geological Survey at the [Hewlett Packard OpenPix Web site](#). Another interesting product can be found at a site maintained by the Fine Arts Museums of San Francisco. Its [ImageBase](#) utilizes yet another promising software program, the Grid Pix software suite, to zoom in on large images, although here it is used for paintings rather than maps (see [Featured Web Site](#)).

Improvements in software, in the storage capacities of servers, in Internet transmission speeds, and in desktop computer processing and display capabilities will no doubt facilitate the use of high-quality digital images of maps over the Internet. However, we are still far from perfection in this area. Downloading times are unacceptable for many purposes, especially for those who must use maps in conjunction with text. For many users the most practical short-range solution may be the purchase of high-quality map images on CD-ROMs or on DVDs. I know that some individuals with an interest in historical maps have been exploring the possibility of marketing such a product, and it may prove economically feasible to sell a variety of maps in this format.

Footnotes:

(1)The technical guidelines for digitization projects at the Library of Congress are available at <http://memory.loc.gov/ammem/formats.html>. The technical guidelines for digitization projects at the National Archives are described in two PDF files that can be downloaded at www.nara.gov/nara/vision/eap/eapspec.html.

(2) Issue No. 12 (1997) of *Meridian* (the journal of the Map and Geography Round Table of the American Library Association) is devoted to the digital imaging of cartographic materials. Footnotes to the articles in that issue provide access to most of the relevant literature.

(3) Flatbed scanners are usually used preferred for imaging projects, since they are least likely to damage fragile materials. The largest flatbed on the market is made by Tangent Color Systems, a division of [Scangraphics Corporation](#). It can copy materials up to 24 x 36" in size, and is being used by the Library of Congress. For materials which are sufficiently durable, roller scanners are also an option. The National Ocean Service is using a color roller scanner made by ANAtech Corporation. According to Curtis Loy, Chief of the Cartographic Technology Program, the large size of many of the Coast Survey Charts necessitated the use of a roller scanner, and he claims that "through judicious use of transparent sleeves to protect fragile documents," no document has been damaged (e-mail communication to the author).

The resolution of digital cameras is continually increasing, and they are becoming a very competitive

option for digital imaging projects. The most powerful camera on the commercial market appears to be the [Phase One Power Phase](#), which has a resolution of 7,000 x 7,000 pixels for medium format cameras (6,000 x 8,400 pixels for large format cameras). Stokes Imaging has recently had a 10,000 x 10,000 pixel camera constructed for its own work, and doubtless cameras of this resolution will soon become more widely available.

It is, of course, also possible to digitize maps by photographing them and then digitizing the resulting negatives or transparencies (as is done in the Icelandic Map Project). Although reasonably good results can be obtained in this way, image resolution and color quality are difficult to control when a photographic intermediate is used, and this method is not recommended if image quality is the most important consideration. On the other hand, processes such as Kodak Photo CD are relatively easy to use and inexpensive, and are fully adequate for many purposes. See Anne R. Kenney and Oya Y. Rieger, "Using Kodak Photo CD Technology for Preservation and Access: A Guide for Librarians, Archivists and Curators." <http://www.library.cornell.edu/preservation/kodak/cover.htm>.

(4) See Steve Gilheany, "The Promise of DVDs for Digital Libraries," *RLG DigiNews*, Vol. 2, No. 3, June 15, 1998. <http://www.rlg.org/preserv/diginews/diginews23.html#technical1>.

(5) Technical information on the Historical Map and Chart Collection was supplied by Curtis Loy, Chief, Cartographic Technology Program, Coast Survey Development Laboratory. Color images are hard to find in this project, but some can be found among the topographic and Civil War maps.

(6) Technical information supplied by Thorsteinn Hallgrímsson and Björn L. Þórðarson of the Icelandic Map Project. Some additional information on this project: The JPEG files used in this project have a CGI file name extension to indicate that they are script generated (users of many software programs will have to change the name extension to JPG to read the images). Because of problems with scanning from photographic intermediates, a digital camera has been purchased for this project, and plans are underway to recapture all digital files directly from the original using the camera.

(7) Both MrSid and Wavelet compression are discussed by Steven Puglia, "Fractal and Wavelet Compression," *RLG DigiNews*, Vol 2, No. 3, June 15, 1998. <http://www.rlg.org/preserv/diginews/diginews23.html#technical2>.

(8) See Kevin Donovan, "The Promise of the Flash Pix Image File Format," *RLG DigiNews*, Vol. 2, No. 2, April 15, 1998. <http://www.rlg.org/preserv/diginews/diginews22.html#FlashPix>.

Technical Feature

Imaging Production Systems at Corbis Corporation

Sabine Sússtrunk, Senior Imaging Researcher, Corbis Corporation

sabines@corbis.com

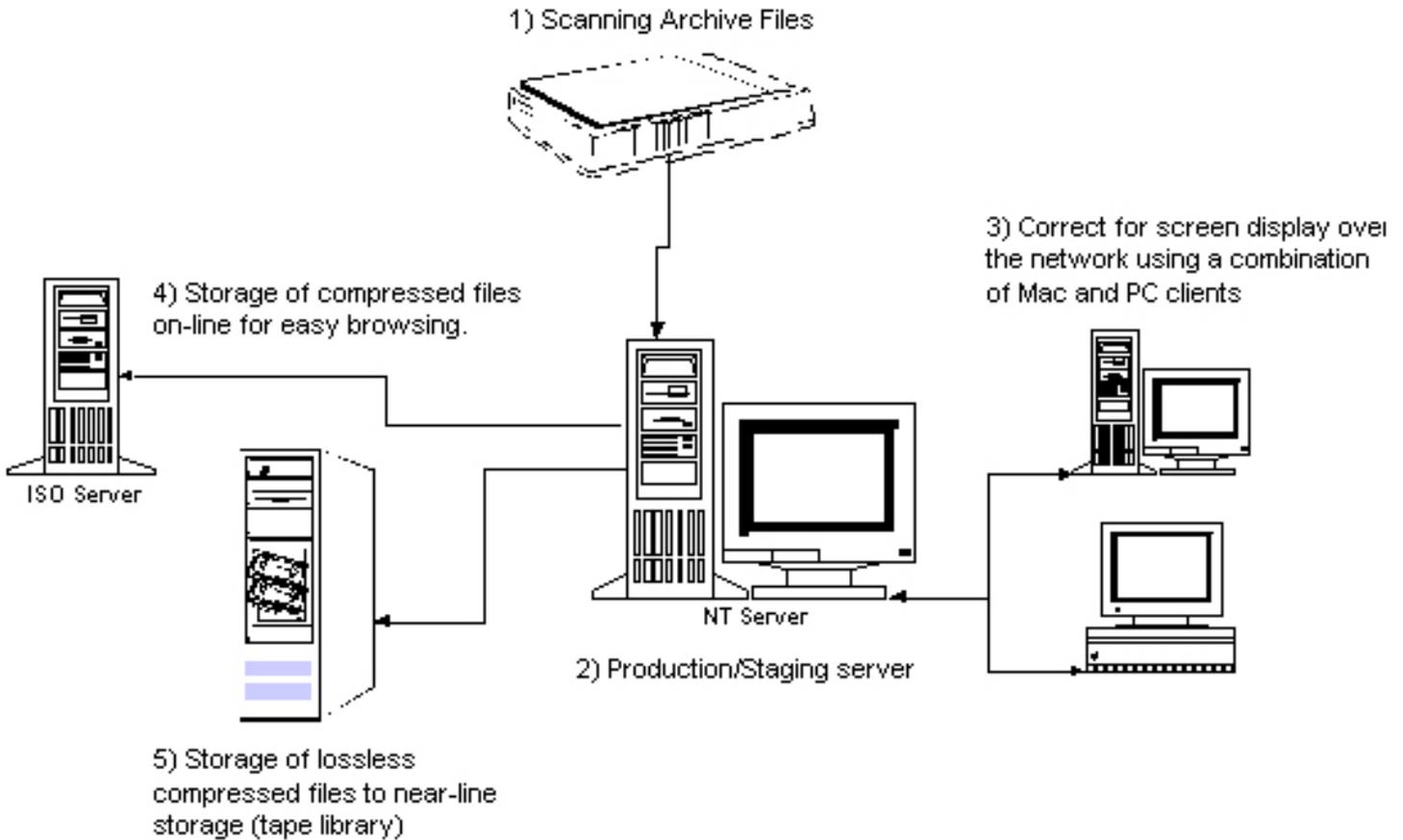
Introduction

Creating a large digital image archive poses a number of technical challenges with regards to capturing, processing, storing, delivering, and managing images. The term "archive" implies that all digitized images are not only optimized for current work flows and imaging devices, but also maintain usability on future, as yet unknown, delivery and output systems. Any production decisions based on currently available technology have to be made with consideration to future image uses and markets.

In the case of a commercial digital archive, the digital image is the product as opposed to the image illustrating a product. Customers are making buying decisions by viewing the images either on screen or on hard copy output, and very often both. This implies that the images have to look "perfect" on any given output, with little or no regard to the tone and color content of the original artwork, which can be a photograph or a painting. In most cases, the customer will never have seen the original, or will only have a memory recollection of it. This application clearly differs from most large imaging initiatives in the museums and libraries community.

How best to scan photographs to create digital facsimiles has been described by Dr. Franziska Frey in the December 15, 1997 issue of RLG's DigiNews. (1) This article will describe how [Corbis](#), a large digital image collection with over 1.3 million high resolution digital images, has addressed the complex issues of digitizing images for an archive with commercial applications as illustrated in Figure 1. The images come from the finest professional photographers, museums, cultural institutions, and public and private collections in the world, and are licensed for one-time use. Corbis and its content providers maintain all copyrights and retain non-exclusive rights to market the images. Images that are used in low-resolution form can be delivered over the Web immediately. High-resolution images can be delivered overnight via CD-ROM, or can be sent via Corbis' FTP (in the US) or ISDN (in Europe) service.

Figure 1: Corbis Archive Production Flow



Corbis Archive Production

1. Scanning Archive Files

Corbis uses high-end Scitex and Heidelberg-Linotype CCD and drum scanners. Although medium priced scanners can currently also produce high quality image scans, high-end scanners such as the Scitex Eversmart or the Heidelberg Topaz offer tools and workflow solutions that allow for higher productivity, such as batch preview and scanning modes. All of the high-end CCD and drum scanners (price range between \$30,000 and \$60,000) on the market today were primarily developed for the pre-press industry. Although all of them allow an RGB or CIE Lab output, many of the tools in the scanning software are optimized for CMYK output, which is not recommended as an archive color space due to its limited color gamut. (2)

1.1 Spatial Resolution

Corbis scans images at the optical resolutions of the scanners used, which vary according to the format of the original. Choosing the correct scan resolution is dependent on the purpose of the archive. Although it would be ideal from an archival point of view to digitize at a level that captures all the information content in an image, in reality the necessary scan resolutions (up to 4,000 pixels per inch for modern film material) are hard to achieve with current CCD scanners for formats larger than 35 mm. The resulting file sizes are huge, up to 960 Mb uncompressed for a 4 x 5 inch original.

It has to be noted that the scanning resolution in pixels (or dots) per inch is not a good indicator of the

resolving power of a scanner or digital camera. The ability of an image capture device to resolve fine detail depends on the whole optical system, including the lens. To get a good indication of the resolution capabilities of a scanning system, the modulation transfer function (MTF) should be measured. (3)

Considering that Corbis is a commercial digital image archive, most images will be reproduced on press. Very rarely will a reproduction exceed 11" x 17", the most common sizes are 8" x 10" or less. The number of pixels per inch required for reproduction depends on the screen frequency of the printing process; usually 1.5 to 2 pixels per halftone dot is recommended. For example, a 150 lines per inch reproduction requires a file of the same magnification with 225 to 300 pixels per inch. Corbis' uncompressed archive file sizes vary between 24 Mb and 60 Mb for an image.

1.2 Image Processing

Very little tone and color correction is applied during scanning. The tone (or gamma) corrections applied are usually to compensate for under- or overexposure, and high or low key scene content. Color corrections are limited to cast removal that was unintentionally introduced in the photographic process. All these corrections use 14-bit or 12-bit controls on the scanners. In practice, these controls change the non-linear quantization of the 14-bit raw scan data to the 8-bit image file data per color channel.

Corbis currently does not save raw scan data, i.e., 14-bit or 12-bit per channel scan data. However, the resulting 24-bit RGB or L*a*b (or Lab) archive image is as close to "raw" as one can practically call a 24-bit scan output. It has to be noted that during all 14-bit to 8-bit quantization of image data on commercial scanning products, the data are interpreted in a non-linear way. Selecting white point, black point, and adding tone and color correction will influence which 8-bit data of the initial linear 14-bit scan signal per channel are retained for the image file. This transform is image dependent and usually not captured during the scanning process, so there is no practical way to recreate raw scan data, or recreate original image data. As Corbis is not in the business of creating digital facsimiles, but excellent image reproductions, that trade-off is seen as an acceptable compromise.

2. Production and Staging Server

The production and staging server, called IVAN, has taken over the function of many manual processes in archive production. After the archive images are downloaded to the server, low resolution image files and histograms are created for initial quality control. Severe image artifacts, such as corrupted scanner look-up tables or clipping in one of the color channels, can therefore be caught early in the production process, and the image can be sent back for rescan. At the same time, the operator ensures that the image identification of the file is the same as the bar code number on the original. Archive image attributes, such as color space, pixel width, pixel height, file format, creation date, etc., are automatically recorded and passed to the image tracking database (ITD). The unprocessed archive images are then moved to the hierarchical storage management (HSM) system for near-line storage (see 5).

After processing, the images are staged for a second quality assurance step. The QA technicians evaluate the quality of processing by thoroughly examining the image file, and comparing it to the original. After QA, three additional low resolution image copies are created (see 4), and the processed images are moved to the HSM system for near-line storage.

3. Image Processing

A resampled copy of the image file (1,280 by 1,024 pixels) is staged on the IVAN production server for Corbis' imaging technicians to access. The technicians, working in a calibrated viewing environment, process the images to a "pleasing color" standard. Pleasing color is defined as global tone corrections to adjust the contrast of the images to the defined output, and to emphasize image specific content. The defined output is a monitor, calibrated to a white point of 6500° Kelvin and a gamma of 1.8. The ambient light is low, resulting in very little viewing glare on the monitor. Color is globally adjusted to balance the image. Additionally, dust and scratches are removed from the image. (4)

The above described image processing steps lead to the creation of an access file. As defined in the archival community, an access file is optimized for a given purpose and output. Corbis intends the access files for monitor viewing. They are used in CorbisViewä, the Corbis image search engine that allows customers to search and preview images and other on-line offerings that can be found on the Corbis Web site (see www.corbis.com). They are also sold to multimedia publishers for on-line and CD-ROM publishing. A resolution of 1,280 by 1,024 pixels offers a full screen image display on a 17 inch, 24-bit enabled monitor.

4. Storage of Compressed Files for On-line Viewing

Corbis stores three resolutions of its access files on-line: a thumbnail, a small preview file with maximum pixel dimensions of 256 pixels (height or width of the image), and a larger preview file with maximum pixel dimensions of 480 by 640 pixels. These files are accessed for previewing by all Corbis applications, internal and external. The images are compressed using JPEG.

5. Near-line Storage

Corbis uses a hierarchical storage management (HSM) system to manage the storage of the high resolution archive files and the 1,280 by 1,024 access image files. The image files are stored in a robotic tape library on DLT tapes that have a capacity of 35 Gb (Gigabytes) of data. The HSM software manages the writing of images to the tapes and the retrieval of images from the tapes. To the user, the image seems to still reside on the image server, and can be retrieved by anyone on the network. The retrieval time of a 30Mb file is about 3 minutes.

The archive images are stored in uncompressed RGB or Lab TIFF format. Lossless compression is achieved by hardware compression when writing the tapes. Corbis does not use a proprietary file format for its archive files. Theoretically, an archive file might not be accessed until the next time the archive is recopied onto the next generation storage medium, which will occur in 7 to 10 years. At that point, source code to interpret the image file format needs to be readily available.

Conclusion

It has been encouraging to see the development of high quality tools destined for digital image applications other than pre-press. However, creating true digital facsimiles that allow the recreation of an original artwork is still very expensive. For commercial digital collections, such as Corbis', archive files represent high-quality image scans that allow for good image reproduction but do not recreate data about the original artwork.

Footnotes:

- (1) Frey, Franziska. "Digital Imaging for Photographic Collections: Foundations for Technical Standards," *RLG DigiNews*, vol. 1, no. 3, December 15, 1997. <http://www.rlg.org/preserv/diginews/diginews3.html#com>.
- (2) Scanners and digital cameras capture colors using the RGB (Red, Green, and Blue) or CIE Lab color model, while printers use the CMYK (Yellow, Magenta, Yellow, and Black) model that is based on the absorption of ink into paper. The Lab model is based on the original color model proposed by Commission Internationale d'Eclairage (CIE), and addresses the problem of the variability of color reproduction that results from the use of different monitors or printing devices. L* is a measurement of the lightness of an object, a* is a measurement of red and green channels, and b* is a measure of yellow and blue channels.
- (3) Williams, Don. "What is an MTF ... and Why Should You Care?," *RLG DigiNews*, vol. 2, no. 1, February 15, 1998. <http://www.rlg.org/preserv/diginews/diginews21.html#technical>.
- (4) The defined output is a monitor, calibrated to a white point of 6500° Kelvin and a gamma of 1.8. Most of Corbis' professional image buyers work in the graphic arts community, which still uses the Macintosh as its preferred platform. The Mac traditionally uses a video gamma of 1.8. On the PC platform, the default gamma is 2.2. A mismatch in monitor gamma makes the "1.8 gamma" image appear darker on a 2.2 gamma monitor. Gamma and color temperature are important factors that can determine the quality of an image display. However, the influence of the viewing conditions can override these two factors. In general, for the human eye to fully adjust to the monitor, the ambient light has to be dimmer than the monitor, and the color temperature of the light source illuminating the room has to be lower (warmer color) than the monitor.
-

Highlighted Web Site

[The Zoom Project](#)

This project is a collaboration between the Fine Arts Museums of San Francisco and the University of California at Berkeley to provide Web access to high resolution images. One of the distinguishing features of the project is the use of the GridPix format, which allows users to scroll within a large image to view portions at different levels of magnification. Similar to Flashpix file format, GridPix uses a tile-based approach, and enables users to access the higher resolution details of an image within one file. Because it is based on server-scripts, the implementation is entirely in HTML, and can be supported by any

graphical Web browser without need for a special software or a plug-in. The Zoom Project Web site further explains the usage of GridPix, and provides examples.

Calendar of Events

[Fourth International Conference on Knowledge Discovery and Data Mining -- KDD-98](#)

August 27-31, 1998

To be held in New York City, this conference will address the science and technology of automated discovery, drawing from the fields of statistics, databases, machine learning, data visualization, high performance computing, knowledge acquisition, and knowledge-based systems.

[OCLC Seminars on Knowledge Access Management Tools and Concepts for Next-Generation Catalogers](#)

Regional Seminar Dates

Minneapolis, Mn., September 14-16, 1998

Philadelphia, Pa., October 14-16, 1998

Birmingham, England, October 19-21, 1998 and October 22-24, 1998

The OCLC Institute is sponsoring an intensive, interactive seminar designed for cataloging leaders. The seminar will provide a comprehensive understanding of cataloging Internet resources--from selection and collection management to cataloging and catalog maintenance.

[Planning for Digitization: Bringing Down Another Barrier To Access](#)

September 17-19, 1998

The Western Council of State Libraries is sponsoring in Denver, Co. a two-day conference designed to inform state libraries and museum staff about issues they should address in planning statewide digitization projects. Topics to be covered include: planning statewide digitization projects, expectations of potential funding sources, and how to find information once it has been digitized.

[The Future of the Humanities in the Digital Age](#)

September 25-28, 1998

The aim of the conference to be held in Bergen, Norway is to educate the participants to better understand the convergence of computing and the humanities. The conference will disseminate information about ongoing projects, and act as an information broker at an international level. For further information contact: futurehum@uib.no.

[The Northeast Document Conservation Center Fall Workshops](#)

The Northeast Document Conservation Center has announced their next workshop series, which includes three that will be of interest to anyone working on digital imaging projects. Further information is available on their Web site or contact: Gay Tracy, tracy@nedcc.org.

Document Imaging and Document Management Course

October 1-3, 1998

To be held in Los Angeles, Ca., this course is designed to teach participants how to manage a document imaging system or digital library. Students will gain an understanding of how document imaging can be used and managed in both small and large scale organizations. Students will learn about the technology of scanning, importing, transmitting, organizing, indexing, storing, protecting, locating, retrieving, viewing, printing, and preserving documents for document imaging systems and digital libraries. For further information contact: SteveGilheany@ArchiveBuilders.com.

Digital Imaging for Libraries and Archives Workshop

October 18-23, 1998

The Cornell University Library Department of Preservation and Conservation will offer a week long workshop on the use of digital imaging technology in libraries and archives, to be held from October 18-23, 1998 in Ithaca, New York. This is the final session of this workshop series, as it will be revised and redesigned during 1999. The next workshop series is tentatively set to begin in October 1999. Enrollment in the October 1998 workshop is limited to sixteen individuals to ensure adequate lab time for participants. There are only a few slots available for this final workshop, and application information can be found at the [workshop Web site](#).

Library and Information Technology Association (LITA) Announces New National Institute

November 16-17, 1998

The Library and Information Technology Association is sponsoring the LITA National Forum 1998: Your Technology Toolkit for 1999 and Beyond, in Oakland, Ca. The LITA Forum will focus on helping both decision makers and practitioners in a variety of libraries and information centers to plan for and manage the technology of the near future. Topics to be covered include Metadata, Unicode, PC and workstation security, authentication and certification of users, technologies for user education, Basic PERL, XML, and Internet2.

OCLC Institute Workshop on Understanding and Using Metadata

This 2 1/2-day seminar on Understanding and Using Metadata will be held six times throughout the Fall. The seminar will provide an intensive theoretical and practical, project-based exploration of metadata and its applications. Topics include understanding emerging Web standards such as XML, RDF, and using Dublin Core metadata in real-world applications. Participants will use the OCLC Institute's Metadata Lab to design, implement, and evaluate a MARC- or Webserver-based resource description system using metadata records for a real-world collection.

Announcements

Final Report of the Nordic Metadata Project Available

The Nordic metadata project final report is now available in HTML format. Nordic co-operation in development of Dublin Core tools will continue in the future. The main goal is the development of tools

needed for creating, gathering, and indexing Dublin Core based metadata.

[The Library of Congress National Digital Library Program Paper Available](#)

"Digital Formats for Content Reproductions," by Carl Fleischhauer, describes the Library's approach to reproducing historical materials, and lists the types of digital formats selected by the Library for this purpose. It includes discussion of formats for pictorial materials, textual materials, maps, sound recordings, and moving-image materials, and is illustrated with examples. Comments or questions about this paper can be directed to: Carl Fleischhauer, [cfl@loc.gov](mailto:cfle@loc.gov).

[Digitisation Forum Online](#)

The Web site from Australia contains over 200 links to sites and papers relevant to the digitisation process. The goal is to keep users up to date with the latest developments, to locate other institutions within the country that are involved in digitisation activities, and to share expertise and experience in this field.

[Library of Congress/ Ameritech National Digital Library Competition](#)

The 1998/99 guidelines are now available on their Web page. The deadline for this year's proposals is November 2, 1998.

[Digital Vault Initiative](#)

UMI has announced a massive effort to digitize its 5.5 billion page archive containing microfilmed documents dating from the first texts printed in English. UMI will provide electronic access to Digital Vault Initiative offerings through online bibliographic databases and both keyword and Boolean searching.

[Internet Library of Early Journals](#)

ILEJ, the Internet Library of Early Journals, a joint project by the Universities of Birmingham, Leeds, Manchester and Oxford, is conducted under the auspices of the eLib (Electronic Libraries) Programme. It has digitised substantial runs of 18th and 19th century journals, and these images are now available on the Internet together with their associated bibliographic data.

[Digital Initiatives Database](#)

The Association of Research Libraries (ARL), in cooperation with the University of Illinois at Chicago (UIC), is developing a database of digital initiative projects underway at its member institutions. The database will include digital projects of all sizes and scope together in one place. The web site is searchable, and lists projects by host institution, technical focus and subject area, and includes fuller description of the project, contact information, and a URL. The database aims eventually to include projects from non-ARL institutions. For more information, contact: Karen H. Zuidema, khz@uic.edu.

FAQs

Question:

We heard a rumor that Kodak is dropping the Kodak Photo CD technology. Is this true?

Answer:

In an attempt to investigate this rumor, we spoke with a number of Eastman Kodak staff members, and reviewed the recent product literature and press releases available on Kodak's Web site. Here is what we found out during this investigation.

Kodak is continuing its support for Kodak Photo CD. The source of this misunderstanding could be Kodak's March 1998 press release announcing the elimination of Kodak Photo CD Imaging Workstation (PIW) from the product line. Kodak Digital Science Conversion System 1200 (CS1200), introduced in June 1996, has replaced PIW. (1) The main force behind this decision seems to be Kodak's interest in providing an open and more flexible environment for its customers and service providers. CS1200 offers many advantages over the PIW:

- CS1200's open system architecture allows conversion of negatives, slides, and prints, whereas PIW is limited to slides and transparencies.
- CS1200 is Macintosh OS-based and can be used for common software applications, whereas PIW is dependent on a SunSPARC system that runs on X Windows with limited interoperability.
- Unlike the PIW that only outputs in Photo CD (Image Pac) and Flashpix formats, the new system provides an array of file format options, including ImagePac, Flashpix, TIFF, and JPEG.
- Overcoming another shortcoming of the PIWs, CS1200 accepts image information from other sources (such as another film scanner, digital camera, etc.) via a Photoshop plug-in.
- CS1200 supports International Color Consortium (ICC) (2) input/output profiles for JPEG and TIFF files, whereas the PIW outputs solely in Kodak Photo YCC color space, requiring special plug-ins for correctly converting color space to the viewing environment (e.g., Acquire for Adobe Photoshop).

More information on Kodak Photo CD services and products can be found at the [Kodak Digital Cameras and Technology site](#).

The changes outlined above provide important benefits for institutions that are interested in converting their collections via Kodak Photo CD technology. For example, a service provider's ability to deliver images in different file formats with ICC profiles is an important added feature. Currently, Web access to ImagePac and Flashpix images require a special plug-in (3), and because of this restriction many institutions convert their Kodak Photo CD images to one of the commonly supported Internet file formats, such as GIF or JPEG. The CS1200 export feature will enable service providers to deliver images in a variety of formats to meet their customers' access and preservation requirements.

In addition to the new features that have been incorporated into the CS1200 products, Kodak announced the availability of a new toolkit for Image Pac files. The toolkit, an application programmer interface (API), allows hardware and software developers to author Image Pac files from their products and take advantage of its multiple resolution structure and platform neutrality.

It is important to note that although Kodak has stopped PIW sales, it will continue to provide service and

support for these products. Kodak also continues to support both ImagePac (PCD) and Flashpix file formats. Currently the Flashpix file format is being developed and promoted by the [Digital Imaging Group](#) that has more than 50 member companies, including Adobe, Canon, Kodak, Fuji, HP, IBM, Intel, Live Picture and Microsoft. Decisions that affect and shape the future of Flashpix will be collectively made by the DIG members.

Kodak continues to support all versions of the Photo CD media, Master, Pro Master, and Portfolio II Photo CDs. The key difference among these options is that Pro Photo CD adds another level of resolution, offering 64 Base images at 4,096 x 6,144 pixel dimensions (the highest resolution that can be achieved via a Master Photo CD is 16 Base, 2,048 x 3,072).

In conclusion, Kodak seems to be committed to continue its support for Kodak Photo CD, and there are no plans to discontinue supporting the ImagePac format for the foreseeable future.

Footnotes:

(1) Both PIW and CS1200 systems include the essential hardware and software required to scan transparencies and slides. Each one includes a Kodak PCD film scanner, data manager, color monitor, CD writer, thermal dye sublimation printer, and system software. CS1200 system also includes a Kodak Digital Print Scanner with 24-bit color and 600 dpi optical resolution.

(2) Further information on the ICC profile can be found in the Technical Feature section of *RLG DigiNews*, vol. 1, no.3, December 15, 1997. <http://www.rlg.org/preserv/diginews/diginews3.html#hardware&software>.

(3) Internet Imaging Protocol (IIP) is developed and promoted by the Digital Imaging Group to allow Web access to Flashpix images without requiring a special plug-in. However, the use of this protocol requires extensive server-based programming to create CGI scripts or server modules. Further information is available at the [Internet Imaging Protocol Web site](#).

RLG News

Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging

Applications to attend the Joint RLG and NPO Preservation Conference on Guidelines for Digital Imaging are now being accepted. This international conference, which carries the overall theme of Guidelines for Digital Imaging, will include speakers from the United States and the United Kingdom, Australia, and Canada. It will be a working conference, designed for those with experience in digital imaging projects who wish to contribute toward the development of international agreement on best practices and uniform approaches to reformatting of library, archival, and museum collections through digitization.

In addition to a keynote address given by Anne R. Kenney (Cornell University) there are three major sessions on:

1. Selection for digital imaging
2. Preparation of materials for digitization
3. Digital image capture

Each of these topics will be addressed by speakers from the US and the UK; conference participants will choose one of the three topics for intensive work to outline common practices and reach consensus on draft guidelines.

Update presentations will be given on:

1. Preservation metadata
2. Digital preservation

And an international panel of experts will provide responses to the conference at its close (speakers include representatives from the UK, Canada, and Australia).

For further details on the program and an on-line application form, see the RLG or UKOLN Web pages at <http://www.rlg.org/preserv/joint/> (from North America and Australia) <http://www.thames.rlg.org/preserv/joint/> (from Europe) or <http://www.ukoln.ac.uk> Mirror site at UKOLN (UK)

Archival Resources: New RLG Service Uses EAD to Enhance and Provide Access to Finding Aids

In the previous two issues of *RLG DigiNews*, the RLG News section has contained information about the use of SGML - and in particular the Encoded Archival Description (EAD) DTD - to improve access to finding aids and the collections they describe. For the past three years, the archival community has worked on the development of a standard approach for presenting finding aids in the electronic environment. To enhance access to these important research tools as they become increasingly available, RLG has created a new service, Archival Resources, which will provide integrated access not only to finding aids, but also to the nearly half million descriptive records of archival collections and items which exist in the RLIN AMC (Archival and Mixed Collections) file.

RLG has worked closely with the archival community in the evolution of the EAD standard, as it is clearly an important new standard for providing access to complex finding aids. It provides a way of retaining their hierarchical structure and improving access to previously hard to find collections. RLG's centralized indexing and display features gives Archival Resources users powerful searching across finding aids, across collections, across institutions, and across oceans. Archival Resources' content comes from major research, academic, and national libraries, archives, and museums in North America and the UK. More

finding aids will continue to be added by RLG's members and like-minded research institutions.

Beginning August 10, you can also explore RLG's new Archival Resources service free of charge, prior to its official release on September 1. For more information about RLG's finding aids program, see the [Archival Resources Web site](http://www.rlg.org/archivalres.html) at www.rlg.org/archivalres.html.

Hotlinks Included in This Issue

Feature Article

[Atlantic Neptune Online](http://mercator.cogs.nssc.ns.ca/neptune.html): <http://mercator.cogs.nssc.ns.ca/neptune.html>

[Hargrett Library at the University of Georgia](http://www.libs.uga.edu/darchive/hargrett/maps/maps.html): <http://www.libs.uga.edu/darchive/hargrett/maps/maps.html>

[Hewlett Packard OpenPix Web site](http://image.hp.com/openpix.html): <http://image.hp.com/openpix.html>

[Historical Map and Chart Collection](http://chartmaker.ncd.noaa.gov/ocs/text/MAP-COLL.HTM): <http://chartmaker.ncd.noaa.gov/ocs/text/MAP-COLL.HTM>

[Icelandic Map Project](http://egla.bok.hi.is/kort/english.html): <http://egla.bok.hi.is/kort/english.html>

[Illinois Aerial Photograph Digital Archive](http://images.grainger.uiuc.edu/airphotos.htm): <http://images.grainger.uiuc.edu/airphotos.htm>

[ImageBase](http://www.thinker.org/imagebase/index.html): <http://www.thinker.org/imagebase/index.html>

[Library of Congress](http://lcweb2.loc.gov/ammem/pmhtml/pandigit.html): <http://lcweb2.loc.gov/ammem/pmhtml/pandigit.html>

[LizardTech](http://www.lizardtech.com): <http://www.lizardtech.com>

[National Archives](http://www.nara.gov/nara/nail.html): <http://www.nara.gov/nara/nail.html>

[Odden's Bookmarks](http://kartoserver.frw.ruu.nl/HTML/staff/oddens/oddens.html): <http://kartoserver.frw.ruu.nl/HTML/staff/oddens/oddens.html>

[Osher Map Library at the University of Southern Maine](http://www.usm.maine.edu/~maps/exhibit2/): <http://www.usm.maine.edu/~maps/exhibit2/>

[Oversized Color Image Project](http://www.columbia.edu/dlc/nysmb): <http://www.columbia.edu/dlc/nysmb>

[Panoramic Map Project at the Library of Congress](http://lcweb2.loc.gov/ammem/pmhtml/panhome.html): <http://lcweb2.loc.gov/ammem/pmhtml/panhome.html>

[Perry-Castenada Map Library at the University of Texas at Austin](http://www.lib.utexas.edu/Libs/PCL/Map_collection/Map_collection.html): http://www.lib.utexas.edu/Libs/PCL/Map_collection/Map_collection.html

[Phase One Power Phase](http://www.phaseone.com): <http://www.phaseone.com>

[San Francisco Bay Area Orthophotoquad Project](http://www.c3.lanl.gov/~bradley/mrsid/bay_area/): http://www.c3.lanl.gov/~bradley/mrsid/bay_area/

[Scangraphics Corporation](http://www.scangraphics.com): <http://www.scangraphics.com>

[SUNY Stony Brook](http://www.sunysb.edu/library/ldmaps.htm): <http://www.sunysb.edu/library/ldmaps.htm>

[TIGER Map Service](http://tiger.census.gov/cgi-bin/mapbrowse-tbl): <http://tiger.census.gov/cgi-bin/mapbrowse-tbl>

[University of Arizona](http://dizzy.library.arizona.edu:80/branches/spc/set/pimeria/intro.html): <http://dizzy.library.arizona.edu:80/branches/spc/set/pimeria/intro.html>

[University of Connecticut](http://magic.lib.uconn.edu/exhibits/ScannedMaps/ScannedMaps.html): <http://magic.lib.uconn.edu/exhibits/ScannedMaps/ScannedMaps.html>

Technical Features

[Corbis](http://www.corbisimages.com/): <http://www.corbisimages.com/>

Highlighted Web Site

[The Zoom Project](http://now.cs.berkeley.edu/Td/zoom.html): <http://now.cs.berkeley.edu/Td/zoom.html>

Calendar of Events

[Digital Imaging for Libraries and Archives Workshop](http://www.library.cornell.edu/preservation/digital.htm): <http://www.library.cornell.edu/preservation/digital.htm>

[Fourth International Conference on Knowledge Discovery and Data Mining -- KDD-98](http://www.aaai.org:80/Conferences/KDD/1998/): <http://www.aaai.org:80/Conferences/KDD/1998/>

[Future of the Humanities in the Digital Age](http://www.futurehum.uib.no/): <http://www.futurehum.uib.no/>

[Library and Information Technology Association \(LITA\) Announces New National Institute](http://www.lita.org/forum98/natfor.htm): <http://www.lita.org/forum98/natfor.htm>

[Northeast Document Conservation Center Fall Workshops](http://www.nedcc.org): <http://www.nedcc.org>

[OCLC Institute Workshop on Understanding and Using Metadata](http://www.oclc.org/institute/metadata.htm): <http://www.oclc.org/institute/metadata.htm>

[OCLC Seminars on Knowledge Access Management Tools and Concepts for Next-Generation Catalogers](http://www.oclc.org/institute/seminar2a.htm): <http://www.oclc.org/institute/seminar2a.htm>

[Planning for Digitization: Bringing Down Another Barrier To Access](http://www.aclin.org/webtele/digit.htm): <http://www.aclin.org/webtele/digit.htm>

Announcements

[Digital Image Database](http://www.arl.org/did/): <http://www.arl.org/did/>

[Digital Vault Initiative](http://www.umi.com/hp/Features/DVault/): <http://www.umi.com/hp/Features/DVault/>

[Digitisation Forum Online](http://www.digitisation.net.au): <http://www.digitisation.net.au>

[Final Report of the Nordic Metadata Project](http://linnea.helsinki.fi/meta/nmfinal.htm): <http://linnea.helsinki.fi/meta/nmfinal.htm>

[Internet Library of Early Journals](http://www.bodley.ox.ac.uk/ilej/): <http://www.bodley.ox.ac.uk/ilej/>

[Library of Congress National Digital Library Program Paper](http://memory.loc.gov/ammem/formats.html): <http://memory.loc.gov/ammem/formats.html>

[Library of Congress/ Ameritech National Digital Library Competition](http://memory.loc.gov/ammem/award): <http://memory.loc.gov/ammem/award>

FAQs

[Digital Imaging Group](http://www.digitalimaging.org/index.html): <http://www.digitalimaging.org/index.html>

[Internet Imaging Protocol Web site](http://www.digitalimaging.org/): <http://www.digitalimaging.org/>

[Kodak Digital Cameras and Technology](http://www.kodak.com/US/en/nav/digital.shtml): <http://www.kodak.com/US/en/nav/digital.shtml>

RLG News

[RLG and NPO Joint Preservation Conference RLG Archival Resources](http://www.rlg.org/archivalres.html): <http://www.rlg.org/archivalres.html>

Publishing Information

RLG DigiNews (ISSN 1093-5371) is a newsletter conceived by the members of the Research Libraries Group's PRESERV community. Funded in part by the [Council on Library and Information Resources](#)

([CLIR](#)), it is available internationally via the [RLG PRESERV](#) Web site (<http://www.rlg.org/preserv/>). It will be published six times in 1998. Materials contained in *RLG DigiNews* are subject to copyright and other proprietary rights. Permission is hereby given for the material in *RLG DigiNews* to be used for research purposes or private study. RLG asks that you observe the following conditions: Please cite the individual author and *RLG DigiNews* (please cite URL of the article) when using the material; please contact Jennifer Hartzell at bl.jlh@rlg.org, RLG Corporate Communications, when citing *RLG DigiNews*.

Any use other than for research or private study of these materials requires prior written authorization from RLG, Inc. and/or the author of the article.

RLG DigiNews is produced for the Research Libraries Group, Inc. (RLG) by the staff of the Department of Preservation and Conservation, Cornell University Library. Co-Editors, Anne R. Kenney and Oya Y. Rieger; Production Editor, Barbara Berger; Associate Editor, Robin Dale (RLG); Technical Support, Allen Quirk.

All links in this issue were confirmed accurate as of August 10, 1998.

Please send your comments and questions to preservation@cornell.edu .

[CONTENTS](#)

[SEARCH](#)

[HOME](#)

[Trademarks, Copyright, & Permissions](#)