




[CONTENTS](#)
[SEARCH](#)
[HOME](#)

Table of Contents


-  **Editors' Note** - Expanded Digital Preservation Focus for *RLG DigiNews*
 - **Feature Articles**
 - **The Persistence of Vision: Images and Imaging at the William Blake Archive** by the Editors and Staff
 - **Economies of Scale: Digitizing the Chicago Daily News** by Matthew Cook
 - **Highlighted Web Site** - Digital Imaging Group
 - **FAQs** - DVD Technology
 - **Calendar of Events**
 - **Announcements**
 -  **RLG News** - Retaining Digital Resources for Tomorrow's Researchers
 - **Hotlinks Included in This Issue**
-

Editors' Note

Expanded Digital Preservation Focus for *RLG DigiNews*

Marie Leveaux, the famous Voodoo Queen, is buried in St. Louis Cemetery No. 1 in New Orleans. Her crypt is always strewn with flowers, coins, and mementos from followers and admirers. On a tour of the cemetery, a guide points out other inhabitants of note - and those less fortunate souls forgotten by family and friends. Their above-ground graves have begun to tilt and sink because the water table is so high in this part of the Mississippi Delta. Left alone, the ground eventually claims these graves. The guide points to a particularly sorry one and says, "no payola." His comment implies that you have to continue to afford to be dead here, and that requires perpetual care.

Maintaining digital resources over time is somewhat analogous to being buried in St. Louis Cemetery No. 1, only the timeframe is much shorter. Both require active intervention to thwart the forces of nature. In the digital world, the forces of nature represent technological change, a lack of institutional will and way, legal impediments, and financial constraints. Preservation in this domain must begin with the assumption that all digital resources will sink into the ground unless they are continuously re-selected for care, and the appropriate methods are used to maintain their presence until the next threat comes along.

In this issue, and every following one, *RLG DigiNews* will focus on some aspect of digital preservation. This issue highlights one of RLG's new initiatives: preserving digital information. In the future, we'll explore the preservation of digital material - images, text, numeric files, sound, and moving images - from many views, examining alternative methods, technical constraints, and cooperative models. We'll present bright, untested ideas; lessons learned (and unlearned!); promising applied research; and maybe even some voodoo. We will examine the technical, social, legal and financial dimensions in articles, FAQs, highlighted Web sites, announcements, and the RLG News section. To make it easier for you to find relevant items, we'll highlight the coverage with our new logo  that incorporates the infinity symbol typically associated with preservation, e.g., denoting the use of permanent/durable paper. But we'll also be trying out some new ideas, so stay tuned. As always, if you have suggestions, comments, or questions, we'd love to hear from you. Just write us at preserve@cornell.edu.

Feature Articles

The Persistence of Vision: Images and Imaging at the William Blake Archive

Editors and Staff, The William Blake Archive

blake@jefferson.village.virginia.edu

Introduction

A free site on the World Wide Web since 1996, the [William Blake Archive](#) was conceived as an international public resource that would provide unified access to major works of visual and literary art that are highly disparate, widely dispersed, and increasingly difficult to gain access to as a result of their value, rarity, and extreme fragility. Currently eight American and British institutions and one major private collector have given the *Archive* permission to include thousands of Blake's images and texts without fees. Negotiations with additional institutions are underway. As of this writing, the *Archive* contains fully searchable and scalable electronic editions of 39 copies of 18 of Blake's 19 illuminated books in the context of full bibliographic information about each work, painstaking diplomatic transcriptions of all texts, detailed descriptions of all images, and extensive bibliographies. When the longest illuminated book, *Jerusalem* (100 plates), appears, the *Archive* will contain at least one copy of every illuminated book printed by Blake and multiple copies of most. It will also feature a searchable new electronic version of David V. Erdman's *Complete Poetry and Prose of William Blake*, the standard printed edition for reference, and detailed hand-lists documenting the complete Blake holdings of the collections from which the contents are drawn. At this point we will then turn our attention to Blake's prodigious work outside the illuminated canon, notably watercolor and tempera paintings, drawings, individual engravings, and commercial book illustrations, as well as manuscripts, letters, and typographic materials. We anticipate that the full *Archive* will contain approximately 3,000 images.

The *Blake Archive* is the result of an ongoing five-year collaboration between its editors (Morris Eaves, Robert N. Essick, and Joseph Viscomi - based at the University of Rochester, the University of California, Riverside, and the University of North Carolina at Chapel Hill, respectively) and the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia. Matthew Kirschenbaum (formerly Project Manager at Virginia, now at the University of Kentucky) serves as Technical Editor. The project has also involved a number of talented graduate students at the editors' institutions (notably Jessica Kem, Kari Kraus, and Andrea Laue). With the efforts of this dispersed group coordinated by daily exchanges on the *Archive's* internal discussion list, we have been able to achieve exceptionally high standards of digital reproduction and electronic editing that are, we believe, models of

their kind. We offer *images* that are more accurate in color, detail, and scale than the finest commercially published photomechanical reproductions and transcribed *texts* that are more faithful to Blake's own than any collected edition has provided. We have applied equally high standards in supplying a wealth of structured contextual information, which includes full and accurate bibliographical details and meticulous descriptions of each image. Finally, users of the *Archive* attain a new degree of access to Blake's work through the combination of text-searching and advanced image-searching tools, including two custom-designed Java applets. The latter allow us to provide a visual concordance to Blake's *oeuvre* - a scholarly resource without precedent in any medium, print or electronic.

Value of Image-Based Archives

The *Blake Archive* is part of an emerging class of electronic projects in the humanities that may be described as "image-based." (1) These projects share a commitment to providing structured access to high-quality electronic reproductions of rare (often unique) primary source materials. Image-based electronic archives also reflect a more generalized shift in editorial theory and practice towards treating texts (of all varieties) as material documents rather than abstract assemblages of words. The result has been renewed attention to such matters as typography, printing methods, bindings, illustrations, and the graphical features of texts, as well as a reluctance to produce so-called "eclectic" editions that borrow from various versions of a work in order to generate an idealized text that accords with the author's putative final intentions. Blake's own title pages, often signed "The Author & Printer W. Blake" are themselves a significant reminder that texts always occupy a physical medium that needs to be accounted for by the editorial process. And while Blake's illuminated books, with their spectacular commingling of word and image, may be taken as unusually dramatic examples of literary works that must be reproduced in forms faithful to their status as material artifacts, scholars in numerous other fields - Dickinson scholars for example, who have found her handwritten (essentially non-pictorial) manuscripts increasingly vital to an understanding of the poet's work-are now undertaking image-based projects of their own.

Figure 1. A typical "Object View" page from the *William Blake Archive*.
Note the centrality of the image in the interface design.

Previous	Index	Next
--------------------------	-----------------------	----------------------

The Book of Thel, copy O, plate 1 (Bentley 2, Erdman ii, Keynes ii)



9292 15.5 x 10.7 cm.

Enlargement	Illustration Information	Transcription
Non-Java	Text Search	Image Search

 [Blake Archive Homepage](#)  [Help](#)

It should be understood that these image-based humanities archives are not strictly definable as digital library projects. Rather, they are an extension of the kind of work traditionally pursued under the aegis of scholarly editing and share with it a rigorous commitment to detail and the relentless critical evaluation of all primary source materials. This in turn bears consequences for the *Archive's* technical practices and procedures. We recognize therefore that our methods may not always be suitable to the needs and production schedules of large-scale digital library projects; nonetheless, we believe that our focus on images, and our thinking about the role of images in humanities research, is relevant to communities beyond the *Archive's* immediate audiences in art history and literary studies.

Digital Image Creation

Digital images are scanned from three types of source media: 4x5-inch transparencies (our preferred and most often used reproductive source) and 8x10-inch transparencies (both of which include color bars and gray scales to ensure color fidelity), and 35mm slides. In most cases, the transparencies have been verified for color accuracy against the original artifact by the editor responsible for color-correcting the digital image, and in all cases by the photographer. Transparencies were initially scanned on a [Microtek](#) Scanmaker III, a flatbed scanner with a transparent media adapter. Since May of 1998, they have been scanned on a Microtek Scanmaker V, which, instead of the adapter

built into the lid, uses a separate drawer to hold 4x5-inch film inside the scanner's body, where it is scanned using Microtek's EDIT (Emulsion Direct Imaging Technology) system (Figure 2).

Figure 2. A project assistant working with 4x5-inch transparencies on the Microtek Scanmaker V.



Scanning transparencies directly and not through glass creates a sharper scan (one less layer to absorb light), eliminates Newton rings (circular impressions), and reduces dust. Slides (which have only rarely been used in the *Archive*) are scanned on a [Nikon](#) LS-3510AF Slide Scanner and a Microtek 35t Plus Scanner. Current versions of Microtek's ScanWizard software (3.1.2) are used with the Microtek scanners; the Nikon runs version 4.5.1 of its supporting plug-in. The Microtek III flatbed and the Nikon slide scanner are attached to a Macintosh PowerPC; the Microtek Scanmaker V is attached to a Macintosh G3; the Microtek 35t Plus is attached to a Macintosh 8600 PowerPC. All three Macs run OS 8.1.

The baseline standard for all scanned images to date is 24-bit color and a resolution of 300 dpi. All scanned images are scaled against the source dimensions of the original artifact so as to display at true-size on a monitor with a 100-dpi screen resolution. (Recognizing that most users access the *Archive* with a lower screen resolution, we have programmed a Java applet to resize images on the fly - see below.) We have found that these settings produce excellent enlargements and suitably sized inline images. Nonetheless, we have begun to scan at 600 dpi now that our hardware (the Macintosh G3 is equipped with 384 Mb of RAM and 9 Gb of internal storage and another 5 Gb external) can better accommodate the resultant larger file sizes, and we have also changed the type of image file we color-correct. Initially, we saved our raw 300-dpi scans as uncompressed TIFFs and as JPEGs, storing the TIFFs on 8mm magnetic Exabyte tape - data which we have since migrated to CD-ROM (in Mac/ISO 9660 hybrid format). Now, by scanning at 600 dpi, we capture more information, including paper tone and texture, but the resulting files are quite large - artifacts a mere 8.5 x 6.5 inches produce files in excess of 100 Mb. Consequently, we store these raw scans on CD-ROM as LZW-compressed TIFFs, which cuts the file size in half. From the original 600-dpi TIFF image, we also derive an LZW-compressed 300-dpi TIFF, which significantly reduces the file size with no noticeable loss of information.

As part of the scanning process for each image, a project assistant has also completed a form known as an Image Production (IP) record (Figure 3). The IP records contain detailed technical data about the creation of the digital file for each image. These records are retained in hard copy at the project's office, and also become part of the Image Information record that is inserted into each image as metadata.

Figure 3. A hard copy Image Production record.

Image Production Record Date _____ Book copy: plate 10001E / 1A1/P-1

	Slide/Transparency Information	Comments
Film stock type	Kodak Ektachrome 100 EBN	
Generation	1st	
Originating Institution	FITZWILLIAM MUSEUM [T 3080]	
Date Slide/Transparency Produced	1998	
Film Type	<input checked="" type="checkbox"/> Color Positive <input type="checkbox"/> Color Negative <input type="checkbox"/> B/W Positive <input type="checkbox"/> B/W Negative	
Dimensions of Source Item	7.9 X 11.9 Inches (cm) mm	
Image Production Information		
Date of Scan	8/1/98	
Person Scanning	TONY STABLEY	
Item Scanned	Slide <input checked="" type="checkbox"/> Transparency (4x3) 8x10	
Scanner Hardware	Nikon LS 3510 (Microtek Scanmaker II S)	
Scanner Software	ScanWizard	
Image Filename	slide 1A1.P4.300.TIF	
Scanner Mode	<input checked="" type="checkbox"/> 24 bit color <input type="checkbox"/> 48 bit color <input type="checkbox"/> line art (b&w) <input type="checkbox"/> greyscale (256 shades; 1000s of shades)	
Orientation	<input checked="" type="checkbox"/> Portrait <input type="checkbox"/> Landscape	
Color Correction	<input checked="" type="checkbox"/> On <input type="checkbox"/> Off	
Scanner Setting	Automatic <input checked="" type="checkbox"/> Manual	
Input Width	5.307 Inches Pixels (cm) mm	
Input Height	8.075 Inches Pixels (cm) mm	
Pitch (Nikon)		
Scaling (Microtek)	148.0	
Output Width	7.856 Inches Pixels (cm) mm	
Output Height	11.95 Inches Pixels (cm) mm	
Scan Resolution	300 <input checked="" type="checkbox"/> DPI PPI	
Final Resolution	300 <input checked="" type="checkbox"/> DPI PPI	
File Size	7.0 KB <input checked="" type="checkbox"/> MP <input type="checkbox"/> Locked <input type="checkbox"/> Free	
Compression	Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>	
File Format	<input checked="" type="checkbox"/> Tif <input type="checkbox"/> Gif <input type="checkbox"/> Jpg	
Aspect	Locked <input type="checkbox"/> Free <input checked="" type="checkbox"/>	
Focus	<input checked="" type="checkbox"/> Automatic <input type="checkbox"/> Manual	
Image Cropped	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No	
Rotation		
Contrast	%	
Brightness	%	
Saturation	%	
Exposure	Blue ___ % Red ___ % Green ___ %	
Channels	Blue Green Red Input ___ output ___	
Curves	Blue Green Red Input ___ output ___	
Filters	Blur Sharpen Edge Enhance	

The 300-dpi compressed TIFF images thus generated are each individually color-corrected against the original

transparency or slide by an editor using Adobe Photoshop and calibrated, hooded Radius PressView 17SR and 21SR monitors (Figure 4). We now color-correct the TIFF file because the lossless compression does not discard image data each time changes are made to the file in Photoshop. This color-corrected TIFF image is then saved on CD-ROM. We do no color correction to the master image file.

Figure 4. Color-correcting a 300-dpi TIFF image in Photoshop. Note the color bar on the TIFF, the monitor hood, and the 4x5-inch transparency visible through the magnifying glass on the lightbox.



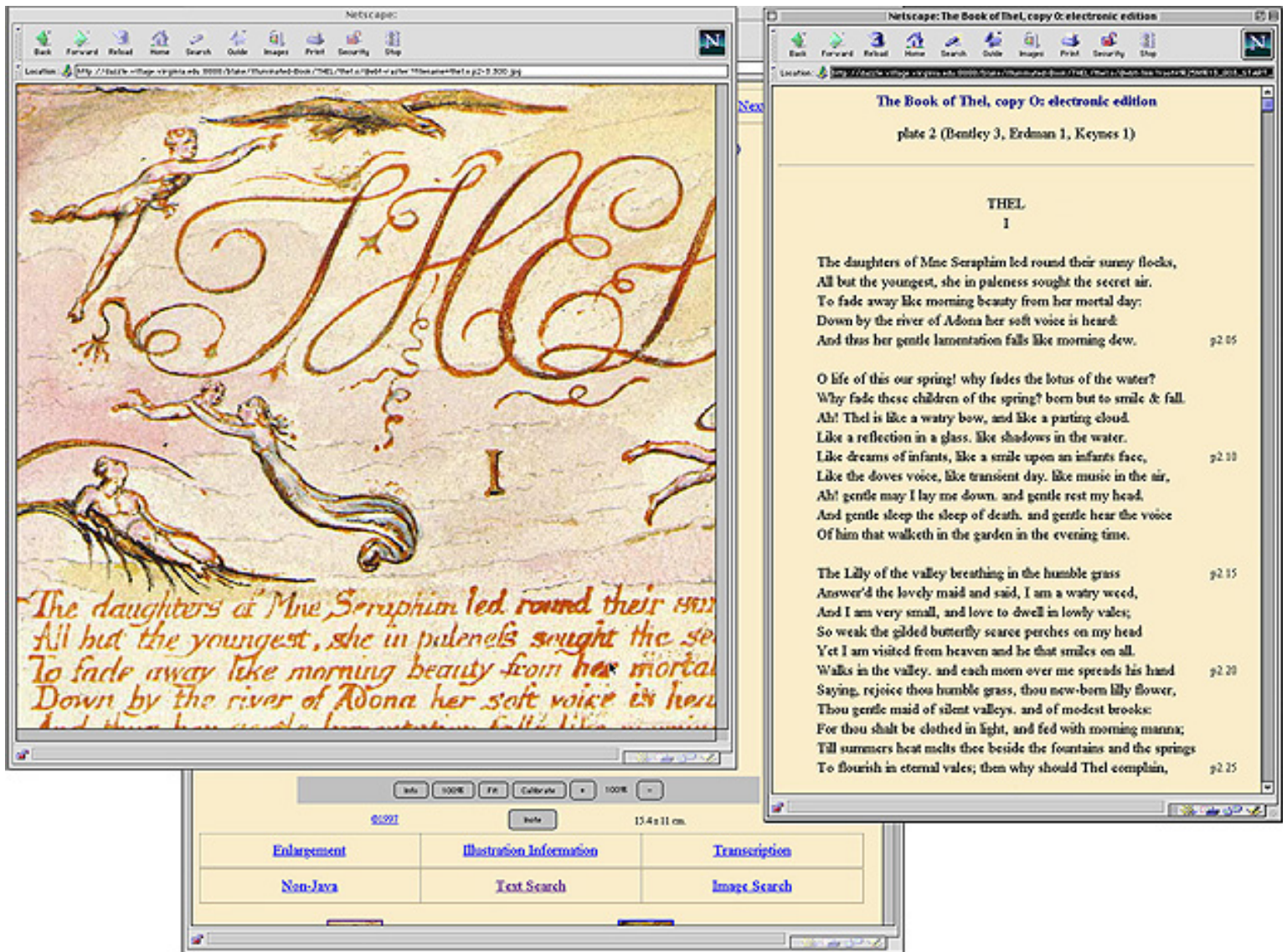
The color-correction process, which takes upwards of thirty minutes - and sometimes as long as several hours - for *each* of the *Archive's* 3,000 images is necessary to bring the color channels of the digital image into alignment with the hues and color tones of the original artifact. This is a key step in establishing the scholarly integrity of the *Archive* because, although we cannot control the color settings on a user's monitor, the color-correction process ensures that each image will match the original artifact when displayed under optimal conditions (which we specify to users). But reproductions can never be perfect, and our images are not intended to be "archival" in the sense sometimes intended - virtual copies that might stand in for destroyed originals. We recognize, however, that if we are going to contribute to the preservation of fragile originals that are easily damaged by handling, we must supply reproductions that are reliable enough for scholarly research. Hence our benchmarks produce images accurate enough to be studied at a level heretofore impossible without access to the originals (Figure 5). In side-by-side comparisons, images in the *Archive* are almost always more faithful to the originals in scale, color, and detail than the best photomechanical images, including the original Blake Trust facsimiles, which were produced through a combination of collotype and hand-painting through stencils. The *Archive's* images create no visual "boundaries" between colors (the result of using stencils), and we have achieved greater color fidelity in some notable instances. For example, several plates in the Blake Trust's facsimile of *The Book of Urizen*, copy G have a reddish hue not found in the originals. In this respect, the images in the *Archive* are truer to the tonality of the originals.

Figure 5. Enlargements of three different impressions all printed from the same plate, showing the variety of coloration (and the importance of color fidelity).



The color-corrected images that we display to our users in the online *Archive* are all served using the JPEG format. A 300-dpi JPEG is derived as the final step of the color-correction process and a 100-dpi JPEG derived from it using ImageMagick, software that enables the batch processing of image files from the UNIX command line. Users are thus presented with an inline image at 100 dpi and have the option to view an enlargement at 300 dpi for the study of details (Figure 6).

Figure 6. A 300-dpi enlargement and diplomatic transcription from *The Book of Thel*, copy O, plate 2.



Metadata Creation

Each and every image in the *Archive* also contains textual metadata constituting its Image Information record. The Image information record combines the technical data collected during the scanning process from the Image Production record with additional bibliographic documentation of the image, as well as information pertaining to provenance, present location, and the contact information for the owning institution. These textual records are, at the most literal level, a part of the *Archive's* image files. Image files are typically considered to be nothing but information about the images themselves (the composition of their pixelated bitmaps, essentially); but in practice, an image file can be the container for several different kinds of information. The *Blake Archive* takes advantage of this by slotting its Image Information records into that portion of the image file reserved for textual metadata (the header). Because the textual content of the Image Information record now becomes a part of the image file itself in such an intimate way, this has the great advantage of allowing the record to travel with the image, even if it is downloaded and detached from the *Archive's* infrastructure. The Image Information record may be viewed using either the "Info" button located on the control panel of the *Archive's* ImageSizer applet (see below) or the Text Display feature of standard software such as X-View. (2)

Specialized Tools

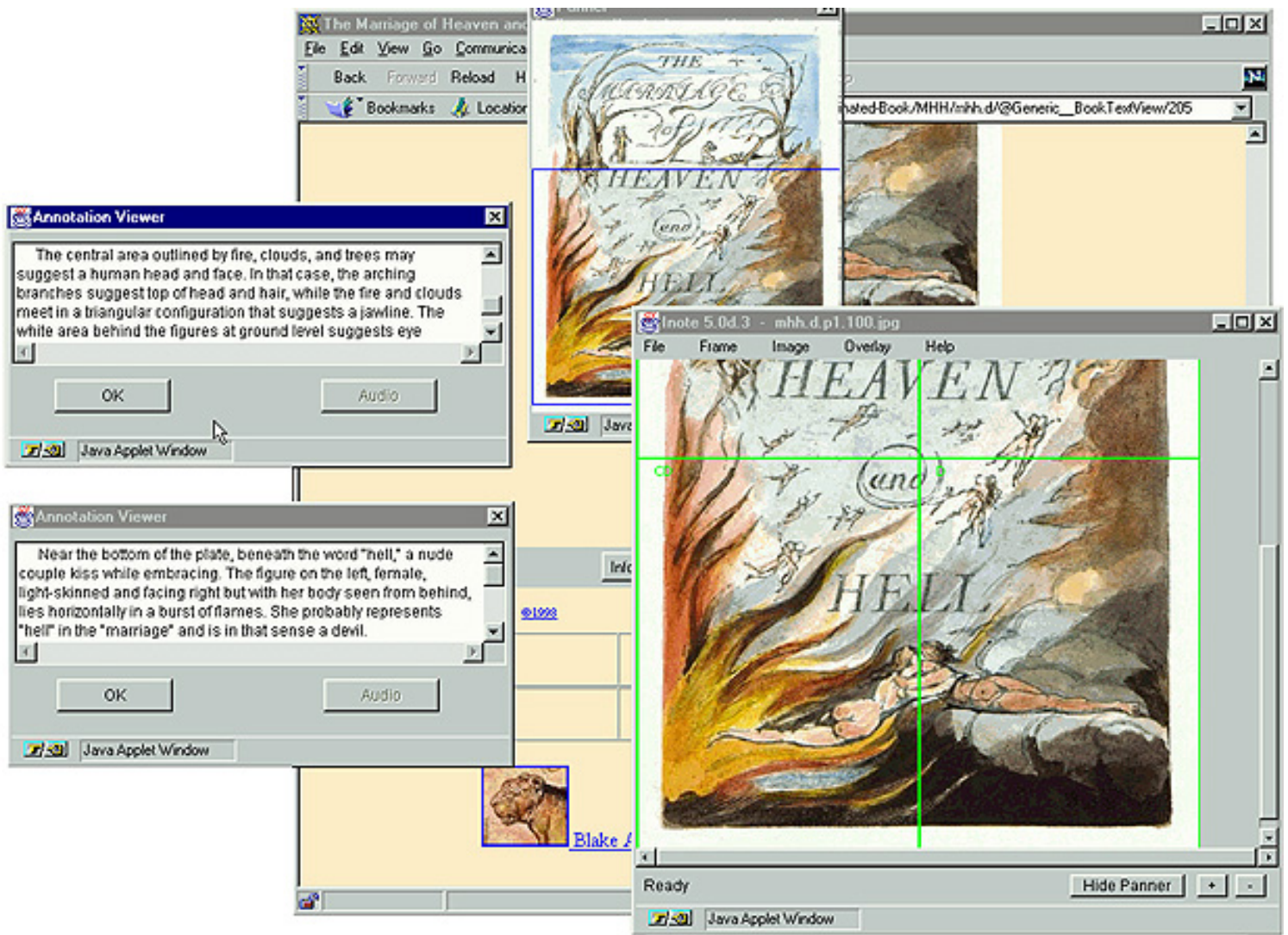
The *Blake Archive* uses two specialized Java applets, developed at IATH in order to support the image-based editing that is fundamental to the project. Both of these applets, Inote and the ImageSizer, should be understood as

computational implementations of the editorial practices governing the design of the *Archive* and its scholarly objectives. Both applets are based on version 1.1 of the Java Development Kit (JDK) and are available for public distribution.

Inote is an image-annotation tool. It permits us to append textual notes ("annotations") to selected regions of a particular image; these annotations are generated directly from the SGML-encoded illustration descriptions prepared by the editors. Inote, which can be invoked from each general description, will bring up the pertinent illustration in a separate window, with access to more elaborate descriptions of individual details within illustrations. The importance of images and their contexts to the project registers powerfully in this Java applet. By means of a location-grid/overlay metaphor, Inote makes it possible to view whole images, components (details) of images, and descriptions of any or all of them (Figure 7). When the user of Inote clicks on any sector of the image, the descriptions of all components in that sector appear. These descriptions are, again, specific to the image being displayed; they are not general descriptions that average (or enumerate) the differences among plates across various copies (instances) of a work. This level of plate-specific description has never been attempted before.

Inote functions most powerfully when used in conjunction with the *Archive's* image searching capabilities, where it can open an image found by the search engine, zoomed to the quadrant of the image containing the object(s) of the search query, with the relevant textual annotation displayed in a separate window. (3) From there, Inote allows the user to resize the image for further study and/or to access additional annotations located in other regions of the image. Inote may also be invoked directly from any of the *Archive's* pages, allowing users to "browse" the annotations created for a given image. In addition, users can download and install their own executable copies of Inote on their personal computers (using a version of the software programmed in the Java Runtime Environment); upon doing so, they may attach annotations of their own making to locally saved copies of an image, for use in either teaching or research. The most recent release of Inote is version 6.0.

Figure 7. A typical Inote session, invoked from the Object View page in the background. The image (*The Marriage of Heaven and Hell*, copy D, plate 1) is displayed in the main Inote window at right, with grid overlay. Inote's navigational Panner, showing the image in its entirety, is also visible. Two image annotations are displayed at left.



The [ImageSizer](#) is a sophisticated image manipulation tool. Its principal function for the *Archive* is to allow users to view Blake's work on their computer screens at its actual physical dimensions. Users may invoke the ImageSizer's calibration applet to set a "cookie" informing the ImageSizer of their own monitor's resolution. Based on this data (recorded in the cookie), all subsequently viewed images will be resized on the fly so as to appear at their true size on the user's screen (Figure 8). If a user returns to the *Archive* at some later date from the same machine, the data stored by the cookie will remain intact, and there will be no need to recalibrate. Users may also set the ImageSizer's calibration applet to deliver images sized at consistent proportions other than true size, for example, at twice normal size (for the study of details). In addition, the ImageSizer allows users to enlarge or reduce the image within its on-screen display area, and to view the textual metadata constituting the Image Information record embedded in each digital image file (Figure 9).

Figure 8. The ImageSizer with its calibration applet. Users hold an actual ruler up against the one on the screen and use the slider bar to align the measurements of both. Screen resolution thus determined, the applet automatically resizes the images to appear at true size on the user's monitor. The applet works by setting a cookie, so calibration need only be performed once for all images in the *Archive*.

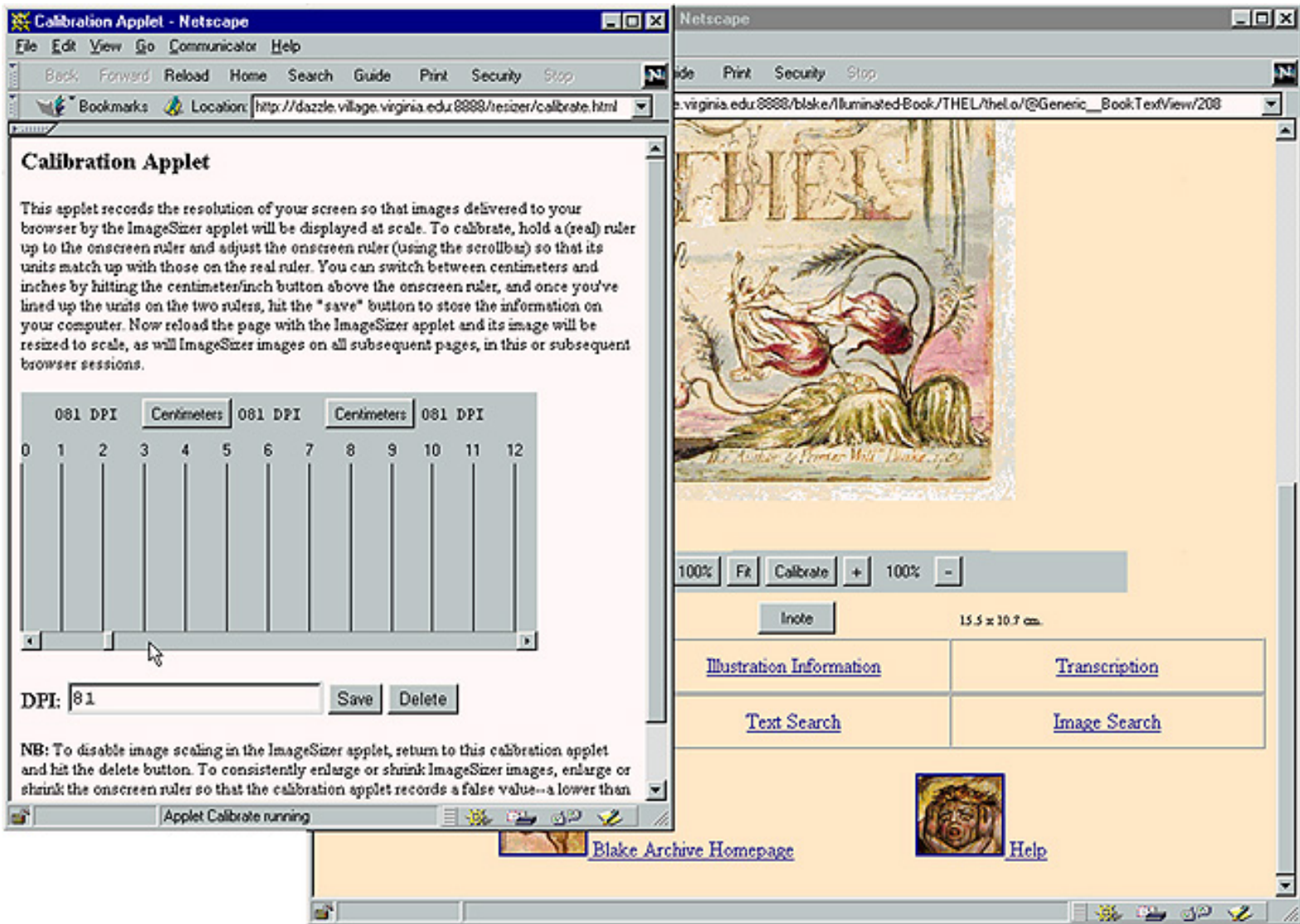
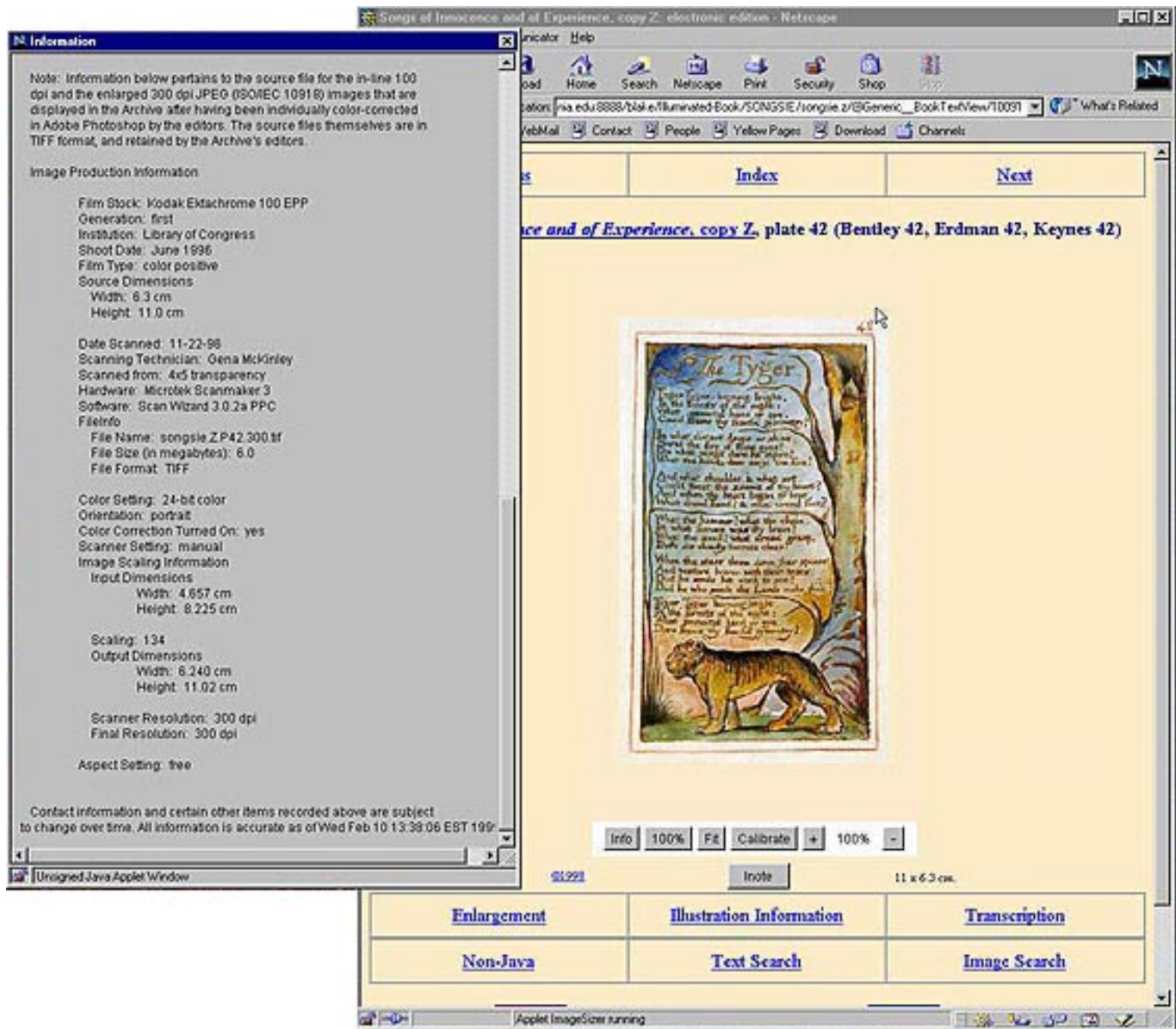


Figure 9. Image Production metadata embedded in the JPEG header, as displayed by the *Archive's* ImageSizer applet.



Future Directions

A few words about the future of images and imaging at the *Blake Archive*. We have been following progress on the [JPEG 2000](#) standard very closely, and have made some preliminary experiments with an image-viewer that uses essentially the same discrete wavelet transform compression technology as the forthcoming JPEG 2000 standard. JPEG 2000 promises a number of advantages that may significantly change the way image-based projects operate. These include lossless image compression, incremental decompression, Inote-like regions of interest, and improved metadata provisions. In particular, we hope that the lossless compression enabled by JPEG 2000 will yield better results than the current JPEG does for the dense line networks characteristic of engravings (interestingly, wavelet technology was originally pioneered by the FBI for compressing large fingerprint databases). Moreover, we expect that the incremental decompression supported by the standard will eventually allow us to keep only one very high quality master copy of an image on our server, and give users the option to retrieve the image (or only a portion of it) at multiple resolutions and levels of detail. Likewise, we have also begun following the International Color Consortium's work on embedded color profiles and anticipate using them to ensure that all users can enjoy the benefits of the painstaking color-correction process we employ. (4)

Finally, we are broadly interested in the future of image-based humanities computing. Although the term "image-based" humanities computing has been in circulation for some time, it appears as if the field is now approaching a

watershed: a number of pioneering projects (many of them begun in the early nineties) whose promise could heretofore be discussed only in speculative terms are now coming to fruition, while new software tools and data standards are poised to redefine the way we create, access, and work with digital images. All of this activity, moreover, is transpiring at a moment when there is an unprecedented level of interest in visual culture and representation in the academic humanities at large. We hope our work at the Blake Archive continues to contribute to these developments. (5)

Footnotes

(1) A very brief and selective list of other image-based projects would include the Rossetti Archive, the Dickinson Electronic Archives, and the Whitman Archive, all based at Virginia's Institute for Advanced Technology in the Humanities <<http://www.iath.virginia.edu/reports.html>>; and the Electronic Beowulf, the result of a collaboration between the University of Kentucky and the British Library <<http://www.uky.edu/~kiernan/eBeowulf/guide.htm>>.

(2) See Matthew Kirschenbaum, "Documenting Digital Images: Textual Meta-data at the Blake Archive," *The Electronic Library*, 16.4 (August 1998): 239-41.

(3) For a detailed discussion of our image searching, see Morris Eaves, Robert N. Essick, Joseph Viscomi, and Matthew Kirschenbaum, "Standards, Methods, and Objectives in the William Blake Archive," *The Wordsworth Circle* 30.3 (Summer 1999): 135-44.

(4) For more on JPEG 2000, see <http://webreview.com/pub/1999/08/13/feature/index3.html>; for more on color calibration, see: <http://webreview.com/pub/1999/08/06/feature/index.html>.

(5) For additional resources on next-generation image-based computing in the humanities, see the LOOKSEE listserv and Web site maintained by Matthew G. Kirschenbaum <<http://www.rch.uky.edu/~mgk/looksee/>>.

Economies of Scale: Digitizing the *Chicago Daily News*

Matthew Cook, Coordinator, Rights and Reproductions, Chicago Historical Society
cook@chicagohistory.org

In August of 1998, the Chicago Historical Society (CHS), the New-York Historical Society, and the Library of Congress were awarded \$520,000 from The Andrew W. Mellon Foundation for a three-year cooperative digitization project. The project's goal is to digitize and catalog for incorporation into the National Digital Library photographic collection materials from the two historical societies. At CHS, the grant is funding the digitization and cataloging of 55,000 glass plate negatives from the *Chicago Daily News* photo morgue, ranging in date from 1900 to 1929. In the course of creating the *Daily News* digital archive, we drew on technical standards and scanning methodologies created by other cultural institutions undertaking similar projects. We considered our project's contribution to be the streamlining of processes, i.e., by digitizing faster and more economically than other projects. This article presents some of the methods that we employed to increase production and reduce costs while meeting the digitization standards established by the library and museum community.

CHS chose to digitize the *Daily News* glass plate collection for both access and preservation reasons. In scanning the glass plates and creating individual MARC records, we are developing a browsable and searchable image database allowing local and remote access to an inaccessible collection (patrons are not allowed to use the glass plates). As for preservation, we believe that by scanning the plates to a high quality standard, we can create prints that would mimic continuous tone photographs and meet the needs of a majority of our users. That is to say, we could create study and publication-quality 8x10-inch and 11x14-inch prints from the scans, enabling us for all intents and purposes to retire the original glass plate negatives, thus ensuring their longevity. In addition, the project provides an opportunity for our staff to inspect, rehouse, and refile each and every plate, helping us to simultaneously identify plates requiring conservation treatment and to create a safe, long-term storage environment in our temperature-controlled negative vault.

Our goal was to scan the original negative *once* in a manner that would meet all of our intended purposes. Similar projects and the literature indicated that for our master or archival scans we would want to create uncompressed TIFF files captured at 8 bits-per-pixel at 600 dpi with resultant files being 7 Mb for the 4x5-inch plates and 12 Mb for the 5x7-inch plates. Some imaging experts suggest capturing 12-bits-per pixel scans. (1) This bit depth was rejected at CHS, as our principal aim was to improve access to the collection. In a field where there are conflicting standards, it is important to establish project goals and create scans that meet those goals. We felt that an 8-bit per pixel scan would meet our goals of access (first) and preservation (second), while simultaneously meeting our production goals.

From our master scan, we create two derivatives: a compressed reference image of approximately 640 pixels on the long end that can be sharpened or otherwise enhanced and saved in JPEG file format; and a thumbnail image of approximately 150 pixels on the long end, sharpened or otherwise enhanced, and saved in the GIF file format. Both sets of derivatives will be served via our Web site and in-house OPAC and searchable via the full-MARC catalog records.

Once we established our scanning criteria, we had to go about determining a production methodology. By talking with colleagues and reading product specifications and pertinent literature, we identified hardware and software that would enable us to create scans per our stated goals. I cannot overemphasize the importance of speaking with - or better yet conducting a site visit with - colleagues who have experience administering digitization projects. Their insights, warnings, and expertise will enrich your and your institution's understanding and contribute to the success of any digitization undertaking.

After examining product information and talking with vendors, we determined that the [AGFA DuoScans](#), a machine already in use by our in-house photo lab and thus possessing a good reputation at CHS, could create digital files meeting our desired specifications. We were happy to discover that DuoScan retails at roughly \$2,000, a modest price, with the downside being that the scanner requires approximately 70 to 90 seconds to complete a scan. Since production was such an issue, we compensated for that time-lag by purchasing two scanning stations that could be operated by one technician. Thus, as one DuoScan captures an image, staff would set up to scan a negative on the adjacent machine. As of this writing, one staff member is generating roughly 25 scans per hour. The price of scanners has dropped dramatically but staff salaries continue to rise. The cost of purchasing two scanners is more than offset by the increased productivity.

Our project scanning staff consists of one full-time project leader and several part-time scanning technicians. For the first seven months of the project, which represented our peak production period, the part-time scanning staff consisted of three people operating the machines six days a week. Since February 1999, we have created 45,000 scans, for an average of 4,000 scans a month. As we begin to wrap up and begin addressing "problem" areas (i.e.,

cracked plates, over-size plates, and some nitrate film), we have scaled back to two part-time scanning technicians and a full-time project leader.

The DuoScans have a pull-out platen on which the transparencies lie. For each machine, we created a frame made of archival matte board, which lies directly on the AGFA platen and into which a technician will place a plate. Unless a plate shows significant damage, we lay the plate emulsion side down to eliminate depth of focus issues. The technician previews the scan using the AGFA proprietary software PhotoLook and at that stage adjusts the histogram reading. This is the only adjustment the technician will make during the capture stage. In attempting to balance quality and production, we felt that by adjusting the histogram, we could rely on the scanner and the PhotoLook software to capture the full tonal range of the negative. By applying contrast adjustment and sharpening only to derivatives, we preserve the raw scan and save image processing time. Because there is no guarantee that any significant number of the 55,000 plates would be printed, we felt that contrast, sharpening, and other digital corrections could be made by experienced staff in the Photo Lab at the time prints are ordered by patrons. Thus, our scanning staff was able to concentrate on production without dedicating a significant amount of time to issues of image manipulation.

After each scan is completed, the technician saves the negative scan based on its *Daily News* negative number (a unique identifier assigned by the newspaper which corresponds to a caption card) and re-sleeves the plate. Because of the repetitive nature of this duty, technicians are scheduled for roughly four-hour shifts. Virtually every other day, the project leader writes to CD-ROM a subset of the previous day's scans to clear the hard drive. We have just recently begun copying the archival scans back to the hard drive and running a batch program in PhotoShop to create the derivatives. Care is taken in creating derivatives to ensure that the images meet a certain quality level when viewed on-screen. The derivatives are then saved to the network for access by CHS catalogers and also written to an archival CD.

To date, we've been able to meet, even exceed, our project goals at a cost lower than we anticipated. In just about a year's time, we have nearly scanned our entire 55,000 glass plate collection, which has allowed us to fold in some other related and hard-to-serve *Daily News* materials, such as over-sized glass plates and a small set of nitrate film. As of this writing, we are creating scans at about \$2.11 per glass plate, with 29% of that comprising equipment costs and 71% going to staff salaries.

But this is just stage one of the project. Stage two, creating intellectual control over the collection, will significantly increase our costs. The creation of full MARC catalog records on a one-to-one basis is time-consuming and expensive. We have just recently begun this stage of the project and expect that it will take about two years. We will shift the scanning staff into support roles for the catalogers as the scanning nears completion so that they can conduct additional research on an as-need basis, thus allowing the catalogers to concentrate on production. The savings in real over projected costs in scanning will enable us to enhance the quality of the cataloging and indexing of the negatives, improving searchability of the materials and hence their value to researchers.

Footnote

(1) See for example, John Stokes, "Imaging Pictorial Collections at the Library of Congress," *RLG DigiNews*, April 15, 1999 < <http://www.rlg.org/preserv/diginews/diginews3-2.html#feature>>.

Highlighted Web Site

Digital Imaging Group (DIG)

The DIG, with approximately 70 member companies, is a consortium engaged in the development and introduction of innovative digital imaging standards and technologies. Among the initiatives of the group are DIG 2000 and DIG 35 (see the "Initiatives" section on the home page). DIG 2000 aims to create file format-related technologies such as a binary container, metadata, and color specification for JPEG 2000 (under development by the JPEG Committee). The focus of the DIG 35 initiative is to define metadata standards for efficient archiving, indexing, and retrieval of image collections.

FAQs

The DVD technology seems poised to take over the market. How does it compare to other formats in terms of cost and capacity?

This issue's FAQ is answered by Tim Au Yeung, Manager of Digital Initiatives, Information Resources Press, University of Calgary.

DVD (which at the time of writing does not have an official expanded name) actually encompasses a number of read only and writeable formats. The writeable DVD formats are the primary focus for this FAQ. An earlier [RLG DigiNews article](#) by Steve Gilheany discussed the promise of DVD for digital libraries.

Writeable DVD comes in three flavors: DVD-RW, DVD-R, and DVD-RAM. DVD-RW has yet to see the light of day and will not be examined at this time. It is this proliferation of standards that has caused confusion among users and uncertainty about adopting DVD as a storage medium of choice. Despite that, the technology holds significant promise in the long term. In the short term, the choice to adopt DVD depends on your requirements. Before examining the relative pros and cons, let's look at the writeable formats.

DVD-R is a write once, read many times format that uses a dye-based substrate as its underlying technology. Currently, only Pioneer manufactures DVD-R drives. The previous generation DVD-R drive cost \$15,000 and was generally intended for mastering DVD-ROM discs. Pioneer has recently announced a second generation DVD-R drive, which costs considerably less (\$5,400) but is still relatively expensive. Two types of media are available; a 3.95 Gb single-sided disc and a newer 4.7 Gb single sided disc. DVD-R discs are compatible with most DVD-ROM drives.

DVD-RAM and DVD-RW are both rewritable formats, utilising a phase-change technology similar to magneto-optical drives. The first entrant in the DVD-RAM field is a drive by Creative Labs and there are a number of manufacturers who have DVD-RAM drives now. DVD-RAM media can store 2.6 Gb per side, for a total of 5.2 Gb on a single disc. The media is available in single and double sided formats. DVD-RAM discs are generally incompatible with DVD-ROM drives although fourth generation (8x) DVD-ROM drives are able to read DVD-RAM discs.

Cost and Capacity

Having done a brief survey of the playing field, it is now possible to compare the formats. Below is a chart comparing major storage media types. All common DVD-RAM and DVD-R formats are included as well as their predecessors, CD-R and CD-RW. For comparison purposes, magnetic disks (hard drives), DLT (digital linear tape), and Jaz and Zip drives are included.

	Cost of Drive	Capacity (in Gb)	Unit Media Cost	Cost per Gb for media	# Needed to Store 20 Gb of Data
CD-R	\$200	.65	\$1.50	\$2.31	31
CD-RW	\$250	.65	\$4	\$6.15	31
DVD-R (single - 4.7 Gb)	\$5,400	4.7	\$45	\$9.57	4
DVD-R (single - 3.95 Gb)⁽¹⁾	\$5,400	3.95	\$40	\$10.13	5
DVD-RAM (single)	\$500	2.6	\$30	\$11.54	8
DVD-RAM (double)	\$500	5.2	\$40	\$7.69	4
DLT-IV	\$2,000+	20	\$70	\$3.50	1
Magnetic disk	\$160 - \$1,000	20	*	\$8 - \$50	1
Jaz	\$300	2	\$100	\$50	10
Zip	\$150	.25	\$15	\$60	80

** for magnetic disks, the cost is based on the drive itself; there are no separate media costs. Magnetic disks have substantial variance in pricing reflecting interface (IDE or SCSI), speed and reliability.*

From a strict cost standpoint, the CD-R comes out quite strongly as the lowest cost format for storage on a per gigabyte basis. When combined with the fact that CD-R drives are among the cheapest around, it makes for a compelling case when examining cost alone. In comparison, DVD-RAM technology costs two to three times more for the drives and three to five times more per megabyte while DVD-R drives cost 20 times more and the per megabyte costs run between three to four times more. This adds up to a substantial difference over the long haul. If you were to store 100 Gb of data, the cost difference between using CD-R and DVD-RAM would be well over \$800.

Preservation Issues

From a longevity stand point, none of the formats has endured sufficient time to get a thorough testing. Most of these formats are less than a decade old and many have been out less than five years. Manufacturers have been claiming 50- to 100-year shelf-life for the CD-R format. Based on this data, some speculate that DVD-RAM and DVD-R formats have analogous life spans. However, these claims need to be taken with a couple of caveats. First, while the shelf life may span that time, handling of any sort can damage the surface of these media types (CD-R/W, DVD-RAM/R) as they have no external protective casing. Minute scratches can render CD-Rs unreadable and DVD-R/RAM with their tighter tracks are even more susceptible to this kind of data loss. Secondly, the shelf life represents laboratory conditions and as stated earlier, have not stood the test of time in the field. CD-R/W and DVD-RAM/R formats are less susceptible to environmental conditions than their magnetic counterparts in DLT, Zip, and Jaz formats, particularly with regards to magnetic fields. In addition, magnetic disk, DLT, Zip, and Jaz formats have

mechanical components than can fail long before the media itself fails. As a final note on longevity, if accidental erasure is a concern then DVD-R and CD-R are the media formats of choice as they are both write once formats that prevent accidental erasure or overwriting of the data.

An issue related to longevity is format compatibility. Even if the media survives for 10 or 15 years, there are few guarantees that there will still be devices available to read the media. Almost no proprietary device has withstood more than 5 years on the market. As users of older tape formats will note, getting the new tape drives to read the older tapes (even when marketed as compatible) can be a considerable challenge. In addition, even if the hardware is capable of reading the older media, the software may not be so equipped. In this area, CD-R and DVD-R do have some advantages. Both formats are designed with conformity to the older, read only standards for CD and DVD. This ensures that generations of CD and DVD that maintain backward compatibility will also maintain compatibility with the CD-R and DVD-R standards. In comparison, CD-RW and DVD-RAM have had incompatibilities with current standards. CD-RW discs can only be read in multi-read CD-ROM drives and second generation DVD-ROM drives. DVD-RAM is only readable on fourth generation DVD-ROM drives, which are only now hitting the market.

Data Access

The situation is less clear when data access is the objective. Manufacturers have been selling CD towers and jukeboxes for a number of years now as a means of providing large quantities of archival data. These mass storage devices come in two flavors; ones allowing near-line storage of data and ones allowing on-line storage of data. On-line storage devices allow for instant access to all the data they store, usually by having as many read devices as their disc capacity. For example, a 66 CD on-line storage device would have 66 CD-ROM drives housed in the controller box. Near-line storage devices allow for instant access to only a small quantity of the data they store at any particular moment but can change rapidly enough that the user can generally wait for the data. An example of this kind of device is a 200 CD tower, which can hold 200 CD-ROMs but may only have 8 CD-ROM drives, allowing instant access to 8 out of the 200 CD-ROMs (with mechanical disc changer system that allows a CD-ROM not in use to be put away and another one brought to the reader drive).

Because the physical configuration between CD-ROM and DVD-ROM/RAM drives are quite similar, the same mechanisms for CD towers and jukeboxes can be used for DVD-R/RAM towers and jukeboxes. This provides a tremendous boon from a storage-for-access stand point. Suddenly a \$20,000 CD tower that housed 66 CD-ROM drives can be converted (by the manufacturer) to house 66 DVD-RAM or DVD-ROM drives for perhaps 10% more in cost. This would give you four times the capacity with DVD-RAM and seven times the capacity with DVD-R than a similar CD-ROM model. One clear advantage these kinds of storage arrays have over more traditional hard-drive based RAIDs is that if the device fails, the data is still intact and can be up and running as soon as the device is replaced. These devices can serve up tremendous quantities of data. Several manufacturers offer models with a 600 disc capacity, translating into a 2.8 terabyte capacity utilising DVD-Rs at the 4.7 Gb capacity range.

Errata:

(1) Clarification note added 28 March 2000: Original text said "(single - 3.95 Mb)."

Calendar of Events

E-Journals Workshop

February 20, 2000, Philadelphia, PA

NISO and NFAIS will jointly sponsor a pre-standardization workshop to explore the need for best practices for electronic publishing and publishing journals on the Web. Problems and issues from the perspectives of primary publishers, secondary publishers, and libraries will be presented.

NINCH Copyright Town Meetings 2000 Series

February 26, 2000 - College Art Association Conference in New York, NY

March 7, 2000 - Triangle Research Library Network in Chapel Hill, NC

April 5, 2000 - Visual Resources Association Conference in San Francisco, CA

May 18, 2000 - American Association of Museums Annual Meeting in Baltimore, MD

With support from the Samuel H. Kress Foundation, the National Initiative for a Networked Cultural Heritage will offer four more Copyright Town Meetings for the cultural community during the year 2000. Issues to be covered include changes in copyright law as it affects working online; fair use and its online future; the status of the public domain; ownership and access of online copyrighted material; distance education; and the development and implementation of institutional and organizational copyright policies and principles.

National Science Foundation Digital Library for Science, Mathematics, Engineering and Technology Education

Letters of Intent Due: March 13, 2000

Building on work supported under the multi-agency Digital Libraries Initiative, this program aims to create a national digital library that will constitute an online network of learning environments and resources for science, mathematics, engineering, and technology (SMET) education at all levels.

Digitising Journals: Conference on Future Strategies for European Libraries

March 13-14, 2000, Copenhagen, Denmark

The conference aims to work out a joint policy of the European national libraries and policy makers on digitizing journals, and hopes to take the first steps towards the establishment of a European JSTOR. The conference is organized by LIBER and Denmark's Electronic Research Library (DEF) in co-operation with the European project DIEPER, the North American JSTOR project, and NORDINFO, the Nordic Council for Scientific Information.

Third International Symposium on Electronic Theses and Dissertations

March 16th-18th, 2000; Preconference: March 15, 2000, University of South Florida, St. Petersburg, FL

This symposium is organized by the NDLTD (Networked Digital Library of Theses and Dissertations), a consortium of research universities committed to improving graduate education by developing digital libraries of theses and dissertations. This conference will serve as a multi-disciplinary forum for graduate deans and their staff, librarians, faculty leaders, and others who are interested in electronic theses and dissertations, digital libraries, and applying new media to scholarship.

Conference on the Economics and Usage of Digital Library Collections

March 23-24, 2000, University of Michigan, Ann Arbor, MI

The Program for Research on the Information Economy (PRIE) and the University Library at the University of Michigan are sponsoring a research conference on the economics and use of digital library collections. The primary audience consists of research scholars, librarians, and publishers with an interest in the challenging opportunities associated with digital publication, distribution, and collection management of scholarly materials.

EVA 2000 Florence: Electronic Imaging & the Visual Arts

March 27-31, 2000, Florence, Italy

Supported by the EC's European Visual Archive Cluster Project, this conference will feature tutorials, workshops, and sessions on leading edge applications in galleries, libraries, museums; new technologies in conservation; museum management and technology; and international initiatives.

Preservation Options in a Digital World: To Film or to Scan

March 30 - April 1, 2000, Brown University, Providence, RI

A Northeast Document Conservation Center (NEDCC) workshop.

Museums and the Web 2000

April 16-19, 2000, Minneapolis, MN

Organized by the Archives & Museum Informatics, MW2000 offers sessions, papers, panels, and mini-workshops on design & development, implementation, evaluation, site promotion, education, societal issues, research, museology, and curation. The Exhibit Hall features hot tools, techniques, and services. Also includes demonstrations of museum Web sites in which participants can meet and question designers and implementers of some of the coolest museums on the Web. Full and half-day workshops precede the conference on April 16.

CIR-2000: The Challenge of Image Retrieval

May 4-5, 2000, Brighton, England

CIR covers all aspects of image and video retrieval from both the UK and overseas. The main themes of CIR-2000 are video asset management, image indexing and metadata, and content-based image retrieval.

IEEE Advances in Digital Libraries 2000 (ADL2000)

May 22-24, 2000, Library of Congress, Washington, DC

IEEE Computer Society Technical Committee on Digital Libraries sponsors ADL2000, with the goal of sharing and disseminating information about important current issues concerning digital library research and technology. This goal will be achieved by means of research papers, invited talks, workshops, and panels involving leading experts, as well as through demonstrations of innovative technologies and prototypes. The conference has the additional goal of indicating the importance of applications of digital library technologies in the public and private sectors of the economy by reporting on and demonstrating available systems.

Moving Theory into Practice: Digital Imaging for Libraries and Archives

First Session: June 19-23, 2000, Cornell University, Ithaca, NY

Offered by the Cornell University Library, Department of Preservation and Conservation, this new workshop series aims to promote critical thinking in the technical realm of digital imaging projects and programs. This week-long workshop will be held four times in 2000 (June 19-23, July 31-August 4, September 25-29, and October 23-27). Each session is limited to 16 individuals. Registration is now open for all four sessions. The workshop is partially funded by the National Endowment for the Humanities.

Digitisation for Cultural Heritage Professionals

July 3-8, 2000, University of Glasgow, Glasgow, Scotland

The Humanities Advanced Technology and Information Institute (HATII) announces the third annual international Digitisation Summer School. This course, designed for archive, library, and museum professionals, delivers skills, principles, and best practice in the digitisation of primary textual and images resources with strong emphasis on interactive seminars and practical exercises.

DRH 2000: Digital Resources for the Humanities

September 10 - 13, 2000, Sheffield, England

Call for proposals. The submission deadline is March 6, 2000.

Call for Hosts: DRH 2001 and 2002, submit proposal by April 8, 2000.

The DRH conferences provide a forum that brings together scholars, librarians, archivists, curators, information scientists, and computing professionals to share ideas and information about the creation, exploitation, management, and preservation of digital resources in the arts and humanities.

Call for Proposals ECDL2000 - The Fourth European Conference on Research and Advanced Technology for Digital Libraries

September 18-20, 2000, Lisbon, Portugal

Deadline for all the proposals is May 1, 2000.

The ECDL series of conferences brings together researchers, industrial members, professionals, user communities, and responsible agents to promote discussion of new emerging issues, requirements, proposals, politics and solutions. ECDL2000 is a co-organisation of the National Library of Portugal, INESC (Instituto de Engenharia de Sistemas e Computadores), and IST (Instituto Superior Técnico - Universidade Técnica de Lisboa). The conference is also supported by ERCIM (the European Research Consortium for Informatics and Mathematics), and the DELOS Network of Excellency (supported by the European Commission).

Announcements

Preservation and Access for Electronic University Records Conference Proceedings Available

In October of 1999, Arizona State University sponsored a conference on Preservation and Access for Electronic University Records. The proceedings are now available on the Web.

Online Copyright Tutorial

The American Library Association, Office for Information Technology Policy, is offering a copyright education course delivered via e-mail. The *Online Copyright Tutorial* is designed for librarians, educators, and researchers.

Improving Access to Early Canadiana: Use of a Digital Collection With Comparisons to Collections of Original Materials and Microfiche

The User-Based Evaluation of the Digital Libraries Research Group at the University of Toronto study is now available on the Web. The focus of the study is on how digital versions of books compare to paper and microfilm copies.

Digitisation of European Cultural Heritage: Products-Principles-Techniques Conference Proceedings Now Available

Held in the Fall of 1999 at Utrecht, this symposium focused on distinctively European digitization projects, and included both plenary sessions and small group demonstrations.

Proceedings from the Telematics for Libraries Meeting Now Available

In November 1999, papers were presented at this meeting and topics such as digital preservation research, and copyright and licensing were discussed.

New Reports from The Library of Congress National Digital Library Program (NDLP)

The NDLP has released two new reports: "The NDLP Writer's Handbook, A guide for the preparation of essays, Web pages and special presentations for American Memory Collections," and "Conservation Implications of Digitization Projects," issued jointly by National Digital Library Program and the Conservation Division of LC.

Garden and Forest: A Journal of Horticulture, Landscape Art, and Forestry

The Preservation Reformatting Division of the Library of Congress has digitized the full set of the first American periodical devoted to horticulture, botany, landscape design and preservation. Included is encoding using the Making of America (MOA) SGML Document Type Definition.

UK Higher Education Joint Information Systems Committee Studies Now Available

More than twenty supporting studies that have been funded as a part of eLib are now all available on the Web. These were commissioned in three thematic areas: studies in digital preservation, evaluation and impact studies, and general studies that include copyright guidelines.

MICROLINK Listserv

The listserv is intended to be a forum for the exchange of issues and ideas related to preservation microfilming and digitization.

European Commission on Preservation and Access Receives EU Grant for Photo Project 'SEPIA'

The ECPA received funding for its project Safeguarding European Photographic Images for Access (SEPIA). Dealing with historic photographic collections, the aim is to promote awareness of the need to preserve photographic collections, and provide training for professionals involved in preservation and digitization of photographic collections.

RLG News

The following article also appears in *RLG News* (no. 50) and is one of several exploring RLG's three new strategic areas. There, the digital preservation initiative is described in this article and in a companion piece written by Peter Graham, University Librarian, Syracuse University. (*RLG News* is a bi-annual print publication. To obtain a copy of the current issue or future issues, send an email request with your postal address to Jennifer.Hartzell@notes.rlg.org).

Retaining Digital Resources for Tomorrow's Researchers

As institutions face an increasing accumulation of electronic research materials to care for and preserve, [RLG's Board of Directors](#) has encouraged the organization to help create a reliable, distributed system of long-term retention repositories, or digital archives, where these materials can be saved into the future.

In the 1980s and early 1990s, RLG brought its major library members together in a collaborative venture to preserve deteriorating print materials. Eight preservation microfilming projects ensured the survival of many thousands of volumes deemed of long-term value to scholarship. Stored in underground vaults, the master negatives will last several hundred years. With materials in digitized form, there is no such guarantee of longevity.

"The reason we are focusing on the long-term retention of digital materials is because of the multi-institutional

character of the problem," says RLG president James Michalko. "The board feels this is something that won't be solved at the single-institution level. How to integrate this activity into their operations is an increasingly urgent challenge for our members. And some of the same mind and skill sets associated with earlier preservation activities are going to have to be brought to bear here."

The issue has ramifications not only for preservation specialists. "The discussion," says RLG program officer Nancy Elkington, "involves collection development people who contribute to decisions about what electronic resources they digitize, subscribe to, or buy; systems people who have to organize and manage access to electronic materials; and directors who worry about how they are going to pay for it all!"

"Until recently," adds program officer Robin Dale, "many libraries could and did rely on compact discs or similar technologies as their storage and delivery mechanisms. But in the past few years, the growing and evolving nature of digital collections has changed this greatly. With the increase in digitization and acquisition of electronic journals and documents, institutions are reaching a critical mass of digital materials they are responsible for. How will they provide continued access to this information? What policies are in place among institutions to support these efforts?"

To find out just how its members were coping, in 1998 RLG commissioned Margaret Hedstrom of the School of Information at the University of Michigan to research and report on [Digital Preservation Needs and Requirements in RLG Member Institutions](#). The bottom line: Among the 54 institutions surveyed, no generally recognized, codified common practices or policies yet exist.

Policies and Practices

The Hedstrom survey informed work now being done by RLG in conjunction with the [Digital Library Foundation \(DLF\)](#), which operates under the umbrella of the [Council on Library and Information Research \(CLIR\)](#). "We are working with DLF members, as well as our own, to help set some policy guidelines for the digital preservation of three different types of resources," explains Anne Van Camp, RLG's manager of Member Initiatives: "electronic records that have never existed in paper form - such as e-mail; electronic publications and journals, many of which are published only electronically - like *RLG DigiNews*, and materials that have been converted into digital form." The [RLG-DLF task force](#) will submit reports at the end of March of 2000.

Defining a "Digital Archive"

There is already a generic reference model to define what a digital archive might look like and how it might function: [OAIS - the Open Archival Information System](#). Developed under the auspices of the Consultative Committee for Space Data Systems,⁽¹⁾ with international input from a broad-ranging set of communities responsible for digital preservation, OAIS is soon to become an international standard. But for RLG's members, responsible for a complex mix of heterogeneous digital files - from business and students' records to electronic journals and images of items in their collections - the generic model needs to be complemented with guidelines specific to research institutions and their particular research resources.

"Our members will need to retain all kinds of digital materials," says Elkington: "print, audio, two- and three-dimensional images, statistical data, and more. What we are trying to do is give them the tools they need in order to budget for and manage the process of moving those digital files - the ones deemed or mandated to have long-term value - from temporary holding places into permanent, durable, and accountable access systems."

Descriptive Metadata

To ensure the universal acceptability of the digital archives envisaged, RLG is talking with other organizations

involved, among them [OCLC - the Online Computer Library Center](#). One common concern is in establishing metadata standards - the descriptive information associated with a digital file. "OCLC has been a leader in developing metadata standards such as the '[Dublin Core](#)' that we know will be very important to all of us concerned with long-term digital preservation in libraries and archives," says Van Camp. "We want to work with OCLC to develop and codify metadata standards that will meet our members' needs for discovery and delivery, as well as long-term retention and access to their own materials."

"A lot of thought has to go into determining common approaches to metadata naming and numbering systems, so that over time separate archives can be at least mutually accessible, if not interoperable," says Dale. "When an institution contracts with a service to archive and provide access to its material, we want to make sure that the work done to prepare those materials for long-term retention is the same across different archives. One would like them all to have more or less the same metadata documentation requirements. By working with OCLC, leaders in CURL's [CEDARS Project \(2\)](#), and the [National Library of Australia](#), we expect to reach consensus on common approaches. These will streamline institutional workflows and improve access to preserved digital objects."

Selection for Long-Term Retention

Beyond the "how to" issues of saving materials long term is the question of "what should be saved?" While individual institutions feel relatively confident in deciding which of their locally digitized collections should be retained into the future, the vast number of digital publications held in multiple locations around the world presents an entirely different set of issues.

Libraries are attempting to work directly with publishers to sort out rights associated with the long-term preservation of electronic journals and other works. National and other legal deposit libraries have taken major strides to pursue these responsibilities in a shared environment, but there is more work to be done. Building on its historic role in defining consortial selection methodologies for preservation microfilming and collection management, RLG will work with a number of partners to develop tools that support and encourage collaborative selection decisions for long-term retention of digital materials.

Partners and Allies

Building the necessary technical and organizational infrastructures to support digital preservation internationally will be a huge undertaking. "The really good news," says Elkington, "is that a host of UK, Australian, and continental European organizations are eager to work with their North American colleagues in sharing leadership, investment, and expertise on this tough set of challenges."

In addition to projects underway at DLF, CLIR, the [Coalition for Networked Information \(CNI\)](#), and RLG, nationally funded work is being done by CURL, by the UK [Arts and Humanities Data Service](#), at the National Library of Australia, and within the [Networked European Deposit Library \(NEDLIB\)](#) group. RLG is involved in several international projects, and staff members are in direct contact with the key players.

"There's lots coming along soon that will contribute usefully to our knowledge base," says Dale. "We're planning an international symposium, as well as some new tools, concept papers, policy statements, and guidelines - all to help move things ahead."

Footnotes

(1) CCSDS is an "international voluntary consensus organization of space agencies and industrial associates

interested in mutually developing standard data handling techniques to support space research, including space science and applications."

(2) CURL is the Consortium of University Research Libraries. CEDARS - "CURL Exemplars in Digital Archives" - is a three-year project focused on strategic, methodological, and practical issues for libraries in preserving digital materials into the future. Part of the UK Electronic Libraries Programme, the project, now in its second year, is led by the universities of Cambridge, Leeds, and Oxford.

Hotlinks Included in This Issue

Feature Articles

AGFA DuoScans: <http://www.agfahome.com/>

ImageSizer: <http://jefferson.village.virginia.edu/inote/download.html>

Inote: <http://www.iath.virginia.edu/inote/>

JPEG 2000: <http://www.jpeg.org/JPEG2000.htm>

Microtek: <http://www.microtek.com/>

Nikon: <http://www.nikon.com/>

William Blake Archive: <http://www.iath.virginia.edu/blake/>

Highlighted Web Site

Digital Imaging Group (DIG): <http://www.digitalimaging.org/>

FAQs

The Promise of DVDs for Digital Libraries: <http://www.rlg.org/preserv/diginews/diginews23.html#technical1>

Calendar of Events

Call for Hosts: DRH 2001 and 2002: <http://www.drh.org.uk/>

Call for Proposals ECDL2000 - The Fourth European Conference on Research and Advanced Technology for Digital Libraries: <http://www.bn.pt/org/agenda/ecdl2000/>

CIR-2000: The Challenge of Image Retrieval: <http://www.la-hq.org.uk/directory/calendar/content4n4.html>

Conference on the Economics and Usage of Digital Library Collections: <http://www.si.umich.edu/PEAK-2000/>

Digitisation for Cultural Heritage Professionals: <http://www.hatii.arts.gla.ac.uk/DigiSS00/>

Digitising Journals : Conference on future strategies for European libraries: <http://www.deflink.dk/english/def.ihtml?fil=digit>

DRH 2000: Digital Resources for the Humanities: <http://www.shef.ac.uk/~drh2000/>

E-Journals Workshop: <http://www.niso.org/ejournalworkshop6.html>

EVA 2000 Florence: Electronic Imaging & the Visual Arts: <http://lci.die.unifi.it/Events/Eva2000/eva2000.html>

IEEE Advances in Digital Libraries 2000 (ADL2000): <http://civic.rutgers.edu/~adl/>

Moving Theory into Practice: Digital Imaging for Libraries and Archives: <http://www.library.cornell.edu/preservation/workshop/>

Museums and the Web 2000: <http://www.archimuse.com/mw2000>

National Science Foundation Digital Library for Science, Mathematics, Engineering and Technology Education: <http://www.nsf.gov/cgi-bin/getpub?nsf0044>

[Ninch Copyright Town Meetings 2000 Series](http://www.ninch.org/copyright/townmeetings/2000.html): <http://www.ninch.org/copyright/townmeetings/2000.html>

[Preservation Options in a Digital World: To Film or to Scan](http://www.nedcc.org/calendar.htm): <http://www.nedcc.org/calendar.htm>

[Third International Symposium on Electronic Theses and Dissertations](http://etd.eng.usf.edu/Conference): <http://etd.eng.usf.edu/Conference>

Announcements

[Digitisation of European Cultural Heritage: Products-Principles-Techniques: Conference Proceedings Now Available](http://www.cs.uu.nl/events/dech1999/): <http://www.cs.uu.nl/events/dech1999/>

[European Commission on Preservation and Access Receives EU Grant for Photo Project 'SEPIA'](http://www.knaw.nl/ecpa/sepia/): <http://www.knaw.nl/ecpa/sepia/>

[Garden and Forest: A Journal of Horticulture, Landscape Art, and Forestry](http://lcweb.loc.gov/preserv/prd/gardfor/gfhome.html): <http://lcweb.loc.gov/preserv/prd/gardfor/gfhome.html>

[Improving Access to Early Canadiana: Use of a Digital Collection With Comparisons to Collections of Original Materials and Microfiche](http://www.fis.utoronto.ca/research/programs/digital/frame.htm): <http://www.fis.utoronto.ca/research/programs/digital/frame.htm>

[MICROLINK Listserv](http://www.slsa.sa.gov.au/microlink/): <http://www.slsa.sa.gov.au/microlink/>

[New Reports from The Library of Congress National Digital Library Program \(NDLP\)](http://memory.loc.gov/ammem/ftfiles.html): <http://memory.loc.gov/ammem/ftfiles.html>

[Online Copyright Tutorial](http://www.ala.org/washoff/alawon/alwn9002.html): <http://www.ala.org/washoff/alawon/alwn9002.html>

[Preservation and Access for Electronic University Records Conference Proceedings](http://www.asu.edu/it/events/ecure/): <http://www.asu.edu/it/events/ecure/>

[Proceedings from the Telematics for Libraries Meeting](http://www.echo.lu/libraries/events/FP4CE/proceed.html): <http://www.echo.lu/libraries/events/FP4CE/proceed.html>

[UK Higher Education Joint Information Systems Committee Studies](http://www.ukoln.ac.uk/services/elib/papers/supporting/): <http://www.ukoln.ac.uk/services/elib/papers/supporting/>

RLG News

[Arts and Humanities Data Service](http://ahds.ac.uk/index.html): <http://ahds.ac.uk/index.html>

[CEDARS Project](http://www.leeds.ac.uk/cedars/index.htm): <http://www.leeds.ac.uk/cedars/index.htm>

[Coalition for Networked Information \(CNI\)](http://www.cni.org): <http://www.cni.org>

[Council on Library and Information Research \(CLIR\)](http://www.clir.org): <http://www.clir.org>

[Digital Library Foundation \(DLF\)](http://www.clir.org/diglib/): <http://www.clir.org/diglib/>

[Digital Preservation Needs and Requirements in RLG Member Institutions](http://www.rlg.org/preserv/digpres.html): <http://www.rlg.org/preserv/digpres.html>

[Dublin Core](http://purl.org/dc/): <http://purl.org/dc/>

[OAIS - the Open Archival Information System](http://www.ccsds.org/RP9905/RP9905.html): <http://www.ccsds.org/RP9905/RP9905.html>

[OCLC - the Online Computer Library Center](http://www.oclc.org/oclc/menu/home1.htm): <http://www.oclc.org/oclc/menu/home1.htm>

[National Library of Australia](http://www.nla.gov.au): <http://www.nla.gov.au>

[Networked European Deposit Library \(NEDLIB\)](http://www.konbib.nl/nedlib/): <http://www.konbib.nl/nedlib/>

[RLG's Board of Directors](http://www.rlg.org/boardbio.html): <http://www.rlg.org/boardbio.html>

[RLG-DLF task force](http://www.rlg.org/preserv/digrlgdlf99.html): <http://www.rlg.org/preserv/digrlgdlf99.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](http://www.rlg.org/preserv/) Web site (<http://www.rlg.org/preserv/>). It will be published six times in 2000. 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 Eden; Associate Editor, Robin Dale (RLG); Technical Researcher, Richard Entlich; Technical Assistant, Allen Quirk.

All links in this issue were confirmed accurate as of February 11, 2000.

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

[CONTENTS](#)

[SEARCH](#)

[HOME](#)

[Trademarks, Copyright, & Permissions](#)