CONTENTS   SEARCH   HOME

Table of Contents

- **Feature Articles**
  - **Digital Imaging Production Services at the Harvard College Library, by Stephen Chapman and William Comstock**
  - **Digital Preservation Conference: Report from York, UK, by Robin Dale and Neil Beagrie**
- **Highlighted Web Site - Digital Library Federation Newsletter**
- **FAQ - Inventories of Digital Imaging Collections**
- **Calendar of Events**
- **Announcements**
- **Hotlinks Included in This Issue**

Feature Articles

Digital Imaging Production Services at the Harvard College Library
Stephen Chapman, Preservation Librarian for Digital Initiatives
stephen_chapman@harvard.edu
and William Comstock , Manager, Digital Imaging and Studio Photography Group
william_comstock@harvard.edu

To meet the growing demand for digitizing capabilities throughout Harvard's libraries, the Harvard College Library (HCL) Preservation & Imaging Department established a digital imaging facility in the summer of 1999. This article describes the rationale for developing in-house production services, and summarizes the planning, configuration, and first-year implementation phases of the program.

Services Offered by the HCL Digital Imaging Group
The mission of the HCL Digital Imaging Group (DIG) is to manage a program for high-quality scanning and digital photography in support of initiatives to create digital collections of permanent research value. To date, DIG has provided digital imaging services exclusively for projects initiated by Harvard librarians and archivists. Projects have ranged from production scanning of photocopies for electronic reserves to color digital photography of trade cards and photographic prints (1). These initial efforts provide a foundation of experience that will help inform the creation of policies, cost analyses, and timetables for future collections reformatting projects.

Planning: Rationale for Establishing an In-House Operation

Harvard's libraries and museums had undertaken several digitization projects (2) before a management group met in late 1998 to consider how to coordinate scanning services for future initiatives. The group discussed Harvard's project experiences and reviewed the published findings of other projects to digitize historic collections. From the outset, HCL sought to establish programmatic services—an infrastructure of staff, systems, and specifications—to digitize printed text, manuscripts, and visual materials.

Initial findings from this analysis suggested that although it is best for the owning library or archives to create descriptive metadata, scanning could be outsourced successfully. However, the management group recognized that significant time must also be budgeted for other activities: materials preparation and transfer, digital image processing and quality control, structural and administrative metadata creation, file management, and storage. This picture of the full digitization workflow ultimately guided HCL to expand the existing services offered by two mature programs, Conservation Services and Preservation & Imaging Services. The goal? Enable curators and project managers to work with a single, flexible organization that could offer a full range of preservation and reformatting services for printed and visual materials.

Two other factors favored the development of an in-house operation. First, there was a priority to digitize special collections, including unique materials, which always present challenges for handling and security. Second, the Harvard University Library had recently initiated a five-year program, entitled Library Digital Initiative (LDI), to build a broad-based infrastructure to select, catalog, deliver, and preserve a wide range of digital collections (3). HCL Preservation & Imaging Services saw an opportunity to make data transfer as efficient as possible by integrating the systems of the local content generators with Harvard's new catalogs and digital repository.

Facility

Widener Library DN-90, home to the Digital Imaging Group, occupies a single studio of approximately 650 square feet. This room is adjacent to the preservation microfilming and photography studios—in operation since 1978 and 1994 respectively—and on the same floor as the 3,400 square foot collections conservation laboratory. Widener "D-level" is restricted to staff-only; the labs are further secured with keycard access.

Jay Willwerth, Head of Imaging Services; Stephen Chapman, Preservation Librarian for Digital Initiatives; and Paul Bellenoit, Director of Facilities Management, were charged with developing DN-90 into a production digital imaging studio. John Howard, Librarian for Information Technology, and the IT department also played a critical role in developing the network and file server infrastructure to meet the demands of managing image files.

Design Criteria

The final fit-out of space and selection of equipment was constrained by a number of requirements, including:

- Accommodating document scanning, digital photography, and image processing/quality control workflows in one studio
- Achieving an optimal balance between handling and production
- Managing color
- Maintaining consistency in processes and products

With these criteria in mind, we consulted colleagues and imaging experts to translate objectives into specifications. In a literature review, we found the Museum of Modern Art's case study to be particularly useful. Its sober assessment of the time needed to set up, troubleshoot, and optimize the many component parts of a studio digital camera helped us set a reasonable schedule to develop the studio (4). David Mathews, then at the Harvard University Art Museums,

Steven Puglia at the National Archives and Records Administration; and Bruce Waterman, then at Corbis Corporation were generous with advice solicited by e-mail and telephone.

Consulting services from two experts proved invaluable. Dr. Franziska Frey, an image scientist at the Image Permanence Institute, provided a comprehensive planning document following a full-day site visit. Her itemization of industry standards relevant to production imaging, including ergonomics, helped us to design individual workstations and to outline workflows. A relatively new standard for viewing images on monitors (soft proofing), *ISO 3664: 2000 Viewing Conditions—Graphic Technology and Photography* (5), guided our selection of all finishes (floor, ceiling, walls, and furniture laminates) and served as the specification for studio lighting (6). Lighting proved to be the most complex task in the design phase of our project. We relied upon an engineer, Jirachai Thongthipaya from Barbizon Lighting Inc., Systems Division, to select the lighting components and to specify the placement of the fixtures so that light levels on both horizontal and vertical surfaces (i.e., monitors) would conform to the standard (7).

Mr. Thongthipaya's considered interpretation of ISO 3664 and our design criteria for the studio led to a creative solution, which turned two 5' x 5' concrete columns from liabilities to advantages. DIG's lighting is divided into multiple zones, with levels independently controlled for each. Materials check-in and document scanning may be conducted in medium-to-bright lighting, digital photography in dim indirect lighting, and critical inspection of images at a very low, set lux (8) level (near dark conditions) as prescribed.

We should note that we view ISO 3664 to be a *partial* solution to the challenge of achieving consistent color and detail reproduction in digital images. Dr. Frey reminded us that image quality, as perceived by a human observer, is influenced by a number of factors. Significant variables include the attributes of the image file, the configuration and age of the viewing device, the observer himself, and the lighting in the ambient environment. The graphic arts standard for soft-proofing gave us the confidence that we could control this last variable in the lab, then replicate the production environment—perhaps at a single workstation in a library, perhaps in an electronic classroom—when critical viewing of still images is essential for research.

Mr. Bellenoit and his staff attended to the complex demands of abatement and retrofit. Overhead pipes, flooring, ceilings, and walls were replaced; two windows were walled over to eliminate external light and dust. Six workstations were included in the floor plan. Cabinets and shelving were designed to store materials of various sizes, supplies, files, and a modest library of technical literature. New electrical (dedicated line) and HVAC (Heating, Ventilating, and Air Conditioning) systems were installed, as were three ceiling-mounted air cleaners. Dust removal was a high priority, particularly for photo digitization projects. Due to the abatement requirements and extensive lead-time required to order lighting, the design and fit-out of the studio, including furnishings, took approximately six months.

Staffing

HCL Imaging Services has been photographing a wide range of library materials for more than 40 years. Experience has taught us that quality depends most on the staff skills. Quality of handling, communications, relationships with subcontractors, turnaround time and price, and, not least, image quality all depend upon skilled management, good training and procedures, and dedicated staff. This is particularly true when studio cameras are used. Being able to light a wide variety of flat and 3D objects effectively requires an experienced and skilled technician, regardless of whether digital or analog photography is used. We are not aware of any technological shortcuts to the judgments and nuanced decision-making required for accurate tone and color reproduction.

Thus, DIG's production services are founded upon highly skilled staff. While the facility was being constructed, staffing was the highest priority in the planning phase. We requested and received approval to post four FTE

positions: manager, photographer, image processing specialist, and scanning technician. The manager, Bill Comstock, joined the staff in the early summer of 1999. Other searches were then conducted under his direction. Due to the strength of the applicant pool for the photographer position—as well as the expectation that special collections scanning projects would soon be proposed in Harvard's LDI Internal Challenge Grant Program—we ended up hiring two photographers, eliminating the need for an image-processing specialist. This brought the total number of photographers in Imaging Services to four. The new photographers joined Stephen Sylvester and Robert Zinck, who were instrumental in establishing the College Library's photography studio in 1994. Areas of specialization and years of experience vary among the four photographers, but all are trained to work in the traditional photography studio, the black-and-white darkroom, and the digital studio, as the work demands. We also hired one scanning technician, with responsibilities defined to cover the full document scanning workflow.

Equipment

We evaluated equipment after addressing staffing requirements. One advantage of this approach is that it is generally easier to obtain the funding to purchase systems that match the skill of the operator than vice versa. Since we would be hiring an experienced manager and photographers, we began the process of evaluating systems knowing that no hardware or software needed to be ruled out due to its complexity.

DIG follows the strategy of purchasing systems just before they are needed for production. Fortunately, collections-level digitization projects generally create sufficient lead-time before scanning needs to begin. (For example, while cataloging is underway, a new type of scanner can be tested, purchased, and installed.) For several reasons, we ruled out microfilm and 35mm scanning for the ramp-up phase. Based upon the projected demand for initial services, the space limitations of the facility, and the number of staff in DIG, we focused our attention on configuring four workstations—one for administrative tasks, two for scanning, and one for image processing and quality control (9). Both scanning workstations were evaluated according to the criteria of handling, flexibility, quality, speed, and integration. Simply put, we wanted each to be capable of doing multiple jobs well.

For document scanning, we selected a Fujitsu M3097 D/G scanner, linked to a standard PC with a 21" monitor. Like other straight-feed duplex flatbed scanners, the Fujitsu can be used for 1-bit or grayscale scanning, in manual- or auto-feed modes. With the capability to scan two sides of a printed text page in a single pass, duplex scanners were determined to be a better match for our applications than higher-priced flatbed scanners with better image processing features. (To date, we have used the Fujitsu scanners—we now have two—only to produce 1-bit images, and, somewhat to our surprise, almost exclusively in the auto-feed mode.)

The second scanning workstation, as we suspected, would turn out to be DIG's single largest expenditure, of money and time. Excepting oversize items, this workstation would be used for all color work and all printed special collections materials. With these applications in mind, as well as the design criteria listed above, we focused our attention on 3,072 x 2,048 area-array digital cameras. Jerome B. Skapof of Academic Imaging Associates, Inc., provided assistance in the multi-month process of evaluating these systems. Although we found the number of camera backs in this product class to be relatively small, we were taken aback by the challenge of integrating copy stand, camera back, camera, lens, lighting, camera software, calibration software, computer, and monitor into a well-tuned production system. After careful review, including evaluation of MTF, color registration, and noise, we selected the Scitex Leaf Volare digital camera back. One consequence of this decision was that DIG would become a two-platform shop (Mac and NT). Good working relationships with our IT department and additional service contracts made this possible, which proved to be advantageous for color management and calibration.

We selected an SGI 320 computer and a Barco Personal Calibrator(r) monitor for the image processing workstation. For start-up, other essential hardware included a GTI Soft Proofer(r) viewing booth, a CD-writer, and, for proof

printing, a black-and-white laser printer and an Epson 3000 color ink jet printer. Key software included specialized applications for color calibration, file renaming, structural metadata creation, CD-ROM production, MD5 checksum generation, script-driven image processing, and image viewing for quality control.

Image processing area including Barco monitor, GTI viewing booth, and DELL workstation with calibrated monitor.

The lessons we learned in selecting and integrating digital imaging equipment will sound familiar to our colleagues who have ventured into production imaging. From the perspective of library IT staff, imaging workstations and applications are highly specialized and require different support resources. While commercial solutions are always preferred, out-of-the-box applications alone are not likely to meet the needs for batch processing and integration with library- or institution-specific catalogs, repositories, and delivery applications. Image and system color management is challenging, but possible, and simplified with high-end equipment and software. Color digital photography, scanning, image processing, file transfer, and storage all demand a robust networking infrastructure.

Year-One Implementation: Workflows and Specifications
Project workflows and imaging specifications are project-specific in Harvard's libraries and archives. Handling requirements, functional requirements for digital reproductions, and budgets vary according to materials, users, and funders/owning libraries respectively. The economics of Harvard's Digital Repository Services—storage will be billed by gigabytes annually—favor approaches that minimize file sizes (without compromising functional requirements). Therefore, the text- and image-scanning workflows described below summarize DIG's project experiences in the first year of operation and should not be construed as across-the-board specifications for Harvard's library digital initiatives.

Text Scanning

Text scanning began in August 1999. This work has largely been tied to two projects: electronic reserves for the Harvard College Library, and the LDI project to digitize annual reports from Harvard and Radcliffe. DIG is now preparing for its fourth consecutive semester of e-reserves. Policies in this program dictate that participating libraries contribute only photocopies to DIG for centralized scanning. As a consequence, turnaround times and production rates are high. Legibility requirements—for page images delivered as image-only PDFs—have been met consistently with 1-bit ITU Group 4 compressed TIFF images, scanned at 300 dpi, or very rarely at 400 dpi, at the discretion of the scanning operator. Auto feed scanning has been used exclusively. For each reserves item, the operator then performs a quality control check of the images, updates the corresponding RV record (Harvard's reserves catalog and management system), and then uploads the TIFF images to a centralized repository, where PDFs are automatically generated.

Because the great majority of the annual reports selected for the LDI project were duplicate copies, curators permitted conservation technicians in HCL Conservation Services to disbind the volumes. Thus, like e-reserves, this project allowed for a page-scanning workflow. We were extremely pleased to find that regardless of embrittlement, all 105,000 pages were auto fed without damage. To be precise, our approach may best be described as "attended auto-feeding." To ensure that no misfeeds or double feeds occurred, the technician often manually fed single pages to the moving rollers. Materials were then passed between two stationary arrays that scanned both page sides in a single pass. At the 600 dpi 1-bit setting we specified for the project, the scanner reliably produced over 650 page images per hour.

We have found that the downstream work steps, however, ultimately define the true rate of production in document scanning. In this project, Mingtao Zhao, DIG's Scanning Technician, not only scanned pages, but then performed the following steps for each report:

- Scripted image "clean up" (despeckling and deskewing)
- Creation of structural metadata (up to four elements per page)
- On-screen inspection of 100% of page images (during metadata production)
- CD-ROM production to send copies of image and metadata files to a subcontractor for OCR and SGML production
- Creation of administrative metadata (formatted in XML) required in Harvard's digital repository
- Transfer of data from DIG to the repository
- Confirmation of receipt of files in the repository

With one technician performing all work steps, as well as tracking anomalies and workflow status in a management database, the baseline rate of production averaged 250 pages per hour.

We have learned that curators welcome services that allow them to "trade" source materials for delivered applications—as opposed to CDs of image files that need to be inspected, processed, and managed. We discovered that using the services of an experienced partner, the University of Michigan Digital Library Production Services, for full-text production worked very well. Finally, due to the initiative of DIG staff, particularly Mingtao Zhao, we found that exploiting the capabilities of Visual Basic, Perl, and Java to automate tasks can increase both the speed and accuracy of all of the production steps following scanning.

Manuscripts and Visual Materials Scanning

Materials in this category submitted to DIG in its first year include 19th-century trade cards (Baker Library); the Hedda Morrison photographs of China (Harvard-Yenching Library), and typescripts from the Nuremberg Military

Tribunal (Harvard Law School Library).

Bringing curators and other stakeholders into the production environment to evaluate test scans early in a project is one of the great advantages of having an in-house operation. All parties benefit: curators have an opportunity to assess all potential risks to the source materials; potential users of the digital surrogates have an opportunity to see which specifications best meet their needs; and DIG has an opportunity to meet the clients and to make sure both parties fully understand project objectives. For collections digitization projects in this category, defining scanning specifications is a group process based upon a review of sample images. These discussions tend to focus upon the preferred size (resolution) of delivery images.

That said, a baseline practice for color imaging is emerging. DIG is indebted to David Remington for bringing some of these techniques to Harvard from the University of California, Berkeley photography studio, where he worked under the direction of Dan Johnston. Mr. Remington joined DIG in the fall of 1999. Guided by ISO 3664 recommendations, DIG calibrates its monitors to 5500°K, Gamma 2.0 for an optimized white point for soft proofing. Archival images are saved as RGB files. Adobe RGB 1998 is used as the working color space when inspecting archival images in Adobe Photoshop. To date, all archival images have been saved as uncompressed 24-bit TIFF files. The KODAK Color Separation Guide and Gray Scale (Q13) has been included in-frame with all manuscripts and images photographed at the Leaf camera. Including this target—a common practice in traditional studio photography—is intended to provide an objective tone and color reference. When photographing a collection of items with shadow and highlight densities that fall within the tonal range characterized by the target, DIG photographers set the tonal response of the digital camera system once per photography session by using the known density values of the target's grayscale patches. By using objective references as guides, we hope to create archival images that, to the extent possible, accurately represent the original items photographed, and not our own subjective, aesthetic biases.

DIG photographer Stephanie Mitchell at the digital camera workstation. An array of the technical targets we used to characterize our systems are evident in the foreground.

DIG has also incorporated the use of system targets for quality control. (Since every scanner and camera introduces biases, generates noise, and fails to reproduce details in certain tonal regions, technical targets document the performance of a system in a given configuration.) DIG uses a variety of targets—Macbeth ColorChecker and ColorChecker DC, Kodak Q-60, and MTF Slanted Edge—to characterize camera setup. As long as the lights and camera position remain stationary, the RGB data in the scanned-target images accurately characterizes the output of that setup for its associated batch of images. DIG's practices are evolving in this area. We make no claim to have finalized an approach to quality control, only to attempting to use objective methods to document image characteristics and imaging system performance.

DIG Photographers, Stephanie Mitchell and David Remington, have also been developing and refining techniques to inspect and process digital images (archival and delivery) more efficiently. DIG routinely produces delivery images from the archival files, but this operation cannot be scripted when curators wish to eliminate targets and excess borders in the thumbnails and "study" images. To facilitate the batch creation of delivery images, photographers save a manually cropped image immediately following photography, while the image is rendered in the camera software. We refer to these images as "production archival images."

Production archival images are sometimes saved in addition to the archival images that include in-frame targets. These cropped, high-resolution images are color corrected using an ICC profile to convert the raw image from the

camera into the Adobe RGB 1998 color space. Derivative images are then sharpened and saved with an embedded Adobe RGB 1998 profile. In creating delivery images, either in batches or individually, we do rely upon subjective aesthetic assessments of images on screen to guide our decisions.

To date, DIG has used UV filtered, cool fluorescent lights at the digital camera for the majority of color work, with strobes used on at least one occasion. Exposure times are relatively short with the area-array camera—approximately 5-10 seconds per color channel. Most flat work is photographed under glass. Conservators have trained DIG in the appropriate handling of printed materials, and have, for example, constructed appropriate cradles to support the photograph albums from the Hedda Morrison project.

Metadata Production

To create structural metadata in text scanning projects, DIG uses a commercial application that functions as a front-end to a Microsoft Access database (10). With the flexibility of Access, one can attempt to output the metadata in the format that is best suited to the delivery application (e.g., XML). Having a technician who can program in Visual Basic greatly increases the flexibility of this application.

DIG creates administrative metadata for all digital still images deposited in the Harvard University Library digital repository (DRS). In order to support file management and preservation services, DRS requires financial, rights, and technical metadata for all deposited objects (11). DIG staff wrote programs in Perl, then Java, to "harvest" much of this metadata automatically from TIFF headers and JPEG files. (Other default metadata are entered manually one time per batch.) The program also generates MD5 checksums for each file, used by DRS programs to validate full receipt of batches.

Future Directions

To accommodate several upcoming projects that require on-site digital photography, DIG has purchased a Nikon D-1 camera. Color flatbed scanners are also being evaluated to meet other needs, such as film scanning. As new projects are proposed, DIG will respond by expanding its analog-to-digital services to meet curators' needs. At the same time, DIG expects to provide digital-to-digital services for images and image collections that are identified for deposit in DRS. These will include inspection of images (existing from earlier projects or "born digital"), creation of derivatives, generation of administrative metadata from file formats in addition to TIFF and JPEG, and batch deposit of images to the repository.

DIG's services are closely aligned with the Harvard College Library Preservation & Imaging Services program as a whole. Planning is underway to provide greater integration among the conservation, materials preparation, studio photography, microfilming, and digitization programs. In addition to satisfying its mission "to create digital collections of permanent research value" for Harvard, DIG will also help the Harvard College Library respond to a growing demand for item-level and over-the-counter digitization services.

Finally, DIG is committed to vetting its practices for image production and image assessment with colleagues. These exchanges, with practitioners at Harvard and other institutions, will not only help DIG refine and stabilize procedures, but will deepen our understanding of which approaches—digital, analog, or both—best meet defined preservation and access requirements.

Footnotes

(1)For brief descriptions of projects, see "Digital Imaging Facility Established," HUL Library Notes (PDF), Nov.1999, p. 5-6. (back)

(2)Selected examples include the Harvard University Art Museums participation in the Museum Educational Site Licensing Project, the Harvard Law School Library's contribution to RLG's "Marriage, Women, and the Law" project, and the Frances Loeb Library, Graduate School of Design project, "American Landscape and Architectural Design, 1850-1920," undertaken in Round One of the Library of Congress/Ameritech competition. (back)

(3) Dale Flecker, "Harvard's Library Digital Initiative: Building a First Generation Digital Library Infrastructure," D-Lib Magazine, Vol. 6, No. 11, November 2000. (back)

(4) Linda Serenson Colet, Kate Keller, and Erik Landsberg, "Digitizing Photographic Collections: A Case Study at the Museum of Modern Art, NY," presented at the Electronic Imaging and the Visual Arts Conference, The Louvre Museum: Paris, September 2, 1997, Spectra 25 (no. 2, 1997): 22-27. (back)

(5) International Organization for Standardization Technical Committee 42 - Photography (in liaison with ISO/TC 130), *ISO 3664:2000 Viewing Conditions - Graphic Technology and Photography*, Second Edition published September 1, 2000. Links to the standard and several related documents are provided at the Library Preservation at Harvard Web Site. (back)

(6) ISO 3664 states that "visible surfaces...should be...colored neutral gray, with a reflectance factor no greater than 0.60." We painted the walls Munsell N 8/. Laminates and ceiling tiles were tested to meet acceptable ranges for reflectance - translated to density of 0.22 or higher, and targeted (not prescribed) CIE Lab values of 95, 100, and 108. (back)

(7) ISO 3664 specifies a single recommendation to achieve consistent viewing conditions to evaluate images on monitors ("soft copy" output). The standard includes parameters for color temperature (5000°K) and intensity of illuminant, reflectance values for surround (any object in the immediate viewing environment), and monitor viewing conditions - including color temperature, white point, and glare. Lighting intensity is specified as a "Reference Ambient Illuminance Level" of 32-64 lx for critical viewing (i.e., quality control) of images on computer monitors. This lighting level is over 10X dimmer than a brightly lit office (approx. 700 lx). Annex D notes that if 24-bit images are displayed in the typical office-viewing environment, "a significant loss in the quality of shadow detail results." (back)

(8) *The Webster's Ninth New Collegiate Dictionary* (Springfield, Mass.: Merriam-Webster, 1991) provides the following definition of lux: a unit of illumination equal to the direct illumination on a surface that is everywhere one meter from a uniform point source of one candle intensity or equal to one lumen per square meter. A scientific definition is provided in Thomas B. Brill, Light: Its Interaction With Art and Antiquities (New York and London: Plenum Press, 1980), 36-39. (back)

(9) Space does not permit enumerating the full technical details of each workstation. Depending upon the interest in this topic, we may post workstation configurations and workflow documents on our Web Site. (back)

(10) See the home page for Informatik, Inc., for descriptions of current products. In 1999-2000, DIG used Informatik's Docuthek program to create tab-delimited structural metadata, as output by Microsoft Access, for corresponding directories of 1-bit TIFF page images. (back)

(11) The DRS Image Metadata Specification is publicly available on the Library Preservation at Harvard "Image Digitization" resources page. This specification provides a data dictionary for the block of the required batch XML to deposit objects in the Harvard University Library digital repository. (back)

Digital Preservation Conference: Report from York, UK

Robin Dale
Program Officer, Research Libraries Group
Robin.Dale@notes.rlg.org
and
Neil Beagrie
Assistant Director DNER, Joint Information Systems Committee UK
Preservation@jisc.ac.uk

Over 150 people from Australia, Europe, and North America convened in York, England, December 7-8, 2000, for an international conference on digital preservation. Sponsored by Cedars (CURL Exemplars in Digital Archives), the Joint Information Systems Committee (JISC), the Research Libraries Group, and OCLC, the conference provided a venue to share, disseminate, and discuss current issues concerning the preservation of digital materials. A key development from the conference was the proposal from JISC and the British Library to develop a Digital Preservation Coalition. In conjunction with the conference, a one-day workshop, Information Infrastructures for Digital Preservation, was held on December 6 to focus on the necessary information infrastructures for preserving digital materials over the long term. Full proceedings from both events will be available on the RLG Web site in the next week, but in the interim, this article serves to summarize and publicize the valuable presentations and discussions that took place.

Information Infrastructures for Digital Preservation

Eight speakers from Austria, Australia, France, Germany, the UK, and the US presented information on their current work in digital preservation metadata to more than 75 participants. This intensive workshop was the first of its kind to focus almost exclusively on *digital preservation metadata*, e.g., metadata that may be used to store technical information that supports preservation decisions and action; to document preservation action taken (migration, emulation, etc.); to record effects of preservation strategies; to ensure the authenticity of digital resources over time; or to note information about collection management and management of rights.(1) Yet despite the seemingly narrow category of metadata, the workshop covered a wide variety of topics and approaches.

Brian Lavoie, OCLC, began the day's presentations by describing the work of the joint OCLC/RLG Working Group on Metadata for Digital Preservation. One of two collaborative efforts to identify and support best practices for the long-term retention of digital objects, this metadata effort is using a consensus-building approach to identify a comprehensive metadata framework to support a broad range of digital preservation activities. Lavoie described a white paper written to launch the work. The white paper, to be made publicly available in January 2001, describes the current thinking and practice on the use of metadata to support digital preservation. Beginning with a definition of

preservation metadata, it progresses to a discussion of high-level requirements for a broadly applicable framework. Highlighting the Open Archival Information System (OAIS) reference model as a common starting point, the paper reviews existing metadata element sets from projects and institutions, which were informed by the OAIS model during their work: Cedars,(2) the National Library of Australia, and NEDLIB.(3) The comparison identified areas of convergence and divergence between those applications and now has laid the groundwork for the OCLC/RLG group to move forward. In the coming months, the group will develop a comprehensive metadata framework, identify the essential preservation metadata elements required to support it, identify and evaluate alternative implementation approaches, and make recommendations for best practices or approaches in implementing preservation metadata.

Following Lavoie's presentation about the OCLC/RLG white paper, an international reaction panel was engaged to provide comments and spark discussion. The panel was comprised of Stephen Chapman, Harvard University; Kelly Russell, The Cedars Project; Colin Webb, the National Library of Australia; and Catherine Lupovici, Bibliotheque Nationale de France, representing the NEDLIB Project. Unanimously, the panel praised the approach taken thus far. Both audience members and panelists, found the comprehensive comparison and mapping between the OAIS and different implementations of it to be an important step toward identifying and documenting a framework, which will be broadly applicable in its support for all types of institutions and all types of digital objects. Lavoie reiterated the fruitfulness of the collaborative process thus far and encouraged workshop participants to send further comment and reaction when the white paper is released.

The remaining set of presentations moved away from the discussion of high-level metadata frameworks and large, collaborative approaches.

Dr. Inka Tappenbeck, State and University Library of Lower Saxony, Goettingen, outlined her work on metadata for terms and conditions and archiving. A subset of work within the three-year CARMEN Project (Content Analysis, Retrieval and Metadata: Effective Networking), CARMEN AP 2/5 is a partnership between the State Library of Lower Saxony and the University of Goettingen, Springer Publishing House, and the Center for Mathematical Sciences, Munich University of Technology. The project objectives are to define the functional requirements for preservation and rights management issues and then to define prototypes of metadata structures. Initial work was with digitized journals and the project combined theory with practice. To identify required metadata elements for the journals and to ensure interoperability between existing sets and the CARMEN AP 2/5 prototype set, existing metadata schemes and best practices such as Dublin Core 1.1, the OAIS, <indecs> , and the Digital Object Identifier (DOI) were analyzed. With a prototype set identified, the set was tested on books and journals digitized by the Goettingen Digitization Centre (GDZ). As CARMEN AP 2/5 continues, the next steps include working with dynamic documents and the development of a workflow for the distributed creation, support, and maintenance of metadata. For more information, contact carmen@www.sub.uni-goettingen.de.

Oya Rieger, Department of Preservation and Conservation, Cornell University Library, discussed the preservation metadata research within Project PRISM, a four-year research initiative between the Library and the Cornell University Computer Science Department. Funded by a DLI2 (Digital Library Initiative) grant, Project Prism (Preservation, Reliability, Interoperability, Security, and Metadata) aims to investigate and develop policies and mechanisms needed for information integrity in the context of a distributed, component-based library architecture. The goal of the preservation metadata component of Project PRISM is to analyze the role of preservation metadata in ensuring information integrity. The metadata framework needs to be effective to form the basis for a preservation service and be extensible to meet the needs of different communities and digital objects. To this end, the library is developing a longevity study to examine and document "vital indicators" that help to monitor the well-being of digital objects. As one starts examing policies and mechanisms required for information integrity, Rieger says, "all roads lead to metadata." The functionality of a preservation service heavily relies on the existence of an enabling and action-

oriented preservation metadata model.

Gunter Muhlberger, EU-Project (European Union), University of Innsbruck, presented information about METAe, the Metadata Engine Project. METAe is a research and development EU-project within the 5th Framework Programme. The key idea of METAe is to systematically extract metadata from the layout as well as from structural and segmental elements of books simultaneously to the digitization process. According to Muhlberger, this is possible because of the predictable nature of the structure of certain books. For hundreds of years, books and journals have had common, recognizable structures and components. Even if a book is printed in a foreign language, a reader could likely distinguish elements such as title pages, indices, chapter headings, etc.—entities that would likely become structural metadata within a digitized version of a book. This theory of digitization and automated metadata extraction is being tested on 17th and 18th century books. The objective of METAe is to develop software that can extract as much metadata as possible from the layout of a book and to transform it into XML structured text. In addition to the structured text, METAe will generate descriptive Dublin Core metadata and the digital facsimile of the document.

In the final presentation, Margaret Byrnes, National Library of Medicine (NLM), discussed the results of the Working Group on Permanence of NLM's Electronic Information, a group convened to recommend establishing permanence levels for resources for which NLM has assumed archiving responsibility. The working group focused on electronic resources that NLM makes available to the public, though the group was aware that its work could provide a model for other publishers of electronic information and also contribute to the development of preservation metadata standards. During its work, the group identified three core categories of permanence for electronic resources and assigned levels of permanence to a number of publications. The permanence levels were then recorded in an existing MARC bibliographic record, or a new MARC record was created to record NLM's permanence commitment. The bibliographic records containing this permanence metadata are to be shared with international bibliographic utilities such as RLIN and OCLC. Users also have access to NLM's permanence metadata through the Web version of the library's public catalog. With permanence metadata proving to be so valuable and well-received, the working group has recommended further work to develop NLM-wide specifications for the format and location of permanence ratings and unique identifiers.

Information Infrastructures: Discussion and Common Threads

While the presentations covered a wide variety of digital preservation metadata and approaches, several common threads emerged throughout the presentations and the day's discussion. These issues were presented and explored in a presentation to the Preservation 2000 conference by Robin Dale.

Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials

A program of fifteen speakers greeted the more than 150 participants of the conference, which was organized into five sessions: Models for distributed digital archives, perspectives of managing national digital collections, practicalities of digital preservation, authenticity and authentication for digital preservation, and "making strides: working together" (collaborative efforts in digital preservation).

Opening Keynote Address

Lynne Brindley, Chief Executive, The British Library, delivered the conference keynote address. In her remarks, she stepped through the issues and challenges associated with digital preservation. First and foremost, she questioned, "Why Preserve?" She responded with a quote from Jeff Rothenberg:

> "The increasing use of digital technology to produce documents, databases, and publications of all kinds has led to an impending crisis, resulting from the absence of available techniques for ensuring that digital information will remain accessible, readable, and usable in the future. Deposit libraries as well as other libraries, archives, government agencies, and organisations must find ways to ensure the longevity of digital artifacts or risk the loss of vast amounts of information and human heritage."

Participants were admonished that "we can no longer rely on benign neglect" as a preservation option in this world of digital preservation. To move forward quickly, she outlined specific issues that must be addressed in the short-term. She paid tribute to the Cedars Project and to the general progress made in the UK in the last five years. She recognized the efforts made on the international front, including the NEDLIB Project, the PANDORA Archive, and the CAMILEON Project. Congratulating those involved, Brindley praised the open and collegial approach taken and encouraged conference participants to take pride in the way the "Medusa-like" agenda of digital preservation is being addressed by those present. Moving on to developments at the British Library, she described the BL's e-strategy, emphasizing the creation of a firm infrastructure and the need to work in close collaboration with peers. Finally, Brindley underscored the urgent need to move beyond projects and to commit to developing a digital preservation agenda at both the national and international levels. A key to this development is the proposal from JISC and the British Library to develop a Digital Preservation Coalition. In closing, Brindley challenged stakeholders to step up to a "national manifesto" for the proposed Digital Preservation Coalition before leaving York and identified eight minimal commitments to be included in such a declaration. The concept was greeted with enthusiastic support and was supported in many presentations and discussions throughout the balance of the conference.

Models for Distributed Digital Archives

Kelly Russell, Project Manager of the Cedars Project, and Vicky Reich, Stanford University Library, each spoke during the session on Models for Distributed Digital Archives.

Kelly Russell provided an overview of the three-year Cedars Project funded by JISC. A multi-institution partnership of the universities of Cambridge, Oxford, and Leeds, the project was established as a response to the need for practical experience in digital preservation. The main objective of the project is to address strategic, methodological, and practical issues and provide guidance in best practice for digital preservation. Russell described the value of using the OAIS reference model not only as an overall reference model for a digital archive, but also as a basis for a shared vocabulary and set of concepts. In just 2.5 years, Cedars has accomplished a great deal, including the development of a working, distributed, demonstrator archive [demonstrations were provided throughout the conference], a specification for metadata for digital preservation, and guidelines for collection managers. To continue building upon the experience and success of the demonstrator, the Cedars project has been granted an additional year of project funding to explore issues and provide recommendations on scaling up the demonstrator archive—perhaps to a national level.

Vicky Reich provided an overview of LOCKSS (Lots of Copies Keep Stuff Safe), a system prototype under development by Stanford University and Sun Microsystems to preserve access to scientific journals published on the Web. Not an archive, LOCKSS is more akin to a "global library system," which focuses on providing persistent access to, and "circulation" of electronic journals. To accomplish this, the LOCKSS system is designed to manage redundant, distributed web caches at "each library for each journal the library wishes to safeguard." Besides the built-in redundancy of the journal caches (Reich estimates that for fairly high use materials, 30-40 copies are needed across different institutions), the system has other persistency features. Caches "pre-fetch" content as it is published so that pages can be preserved even if they are not read. Also, the system continuously validates against other caches, and

can detect and repair failure of caches or files by automatically "requesting" replacement copies from the publisher and other caches. Alpha testing of the prototype is underway and emerging results show LOCKSS to be feasible. Further developments to the system are in process and beta testing will begin in 2001.

Perspectives of Managing National Digital Collections

This session explored three different approaches to managing national digital collections through Helen Shenton, Deputy Director, Collections Management (Preservation) for the British Library, Lex Sijtsma, IT Coordinator for Koninklijke Bibliotheek, and Colin Webb, Director of Preservation at the National Library of Australia.

Helen Shenton's presentation described how the British Library (BL) is addressing the preservation of digital material, in particular how it is working on the preservation component of the Digital Library System (DLS), and how staffing issues are affected by new responsibilities. According to Shenton, the digital preservation responsibilities at the BL include preservation of digital material; conservation and preservation of the physical carrier of the digital object; use of digital technology in the curatorial and conservation decisions; exploration of 'hybrid' surrogacy; and development of policies, strategies, and procedures to underscore all the activities mentioned. Shenton stressed the obvious implications that digital preservation has had for staffing an institution that must deal with 'the old and the new.' She also discussed the benefits of collaborations with other entities, both from within the library and without. It is clear that the BL, and BL Preservation in particular, are in a state of transition. However, with the digital preservation infrastructure beginning to emerge, the BL seems poised to fulfill the library's e-strategy for the future.

Lex Sijtsma addressed the role of the Koninklijke Bibliotheek (KB) in the Networked European Deposit Library (NEDLIB) Project. NEDLIB's goal is to construct the basic infrastructure upon which a networked European deposit library of digital publications can be built. Similar to the Cedars Project, NEDLIB looked to the OAIS reference model as a starting point for process and data modeling efforts. The resulting data model is based on the OAIS, but includes a specific preservation component within it. The project has also completed a workflow model and a series of reports including guidelines for depositing electronic publications, using emulation for long-term preservation, and metadata for long-term preservation. A workshop has been developed to share the results of NEDLIB (Setting Up Deposit Systems for Electronic Publications—dSEPs) and institutions are beginning to implement dSEPs locally. Sijtsma's presentation concluded with a walk-through of a dSEP and a discussion of KB's upcoming implementation of a dSEP—project DNEP—to be operational in 2002.

Colin Webb discussed the National Library of Australia's perspective on managing a national digital collection. Since 1996, the NLA's PANDORA Project (Preserving and Accessing Networked Documentary Resources of Australia), has been archiving selected Australian online publications such as electronic journals, organizational sites, government publications, and ephemera. The NLA also assumes responsibility for the long-term preservation and accessibility of digital materials deemed to be of national interest. To address those needs, the library identified the requirements for robust, high performance systems to manage its present and future digital collections and to support shared access to digital collections in cooperation with other institutions. Along with the holistic selection approach for digital publications, the Library also takes a holistic approach to digital preservation, preferring to differentiate between "archiving" and "long-term preservation." With a national remit to "ensure the viability" of certain materials "in perpetuity," Webb's presentation outlined the different strategies the NLA will employ to accomplish that goal while considering some of the strengths and weaknesses of the NLA's approach.

Practicalities of Digital Preservation

Focusing on the practicalities of digital preservation, this session included presentations from Margaret Jones, The

Arts and Humanities Data Service; Michele Cloonan and Shelby Sanett, University of California, Los Angeles, and Ellis Weinberger, Cambridge University Library

Margaret Jones presented information about an Arts and Humanities Data Service (AHDS) research project on digital preservation management. Because digital preservation lacks practical, widely available tools, the project aimed to provide guidance, including a decision tree and other tools, to assist organizations in managing the digital materials they create and acquire. The main deliverable from the project was *A Workbook for the Preservation Management of Digital Materials*, produced by Jones and Neil Beagrie. A peer review draft of the workbook is currently available from the JISC Digital Preservation Focus webpages.In her presentation, Jones reviewed the feedback provided during the initial period of peer review and detailed how the feedback impacted the final draft of the workbook as well as the upcoming workshops based on the workbook.

Michele Cloonan and Shelby Sanett jointly reported on a study being conducted on behalf of the Preservation Task Force of the InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems). The InterPARES Project is an international research initiative involving national archives, university archives, and various government agencies who are working together to address issues related to the permanent preservation of authentic electronic records. The purposes of the Cloonan and Sanett study were to learn more about testbed projects, and to compare and contrast the "genesis of and perceived effectiveness of" preservation methods currently underway or in development. Issues explored during the study included the existence of formal policies, strategies being employed, and the perceived cost of digital preservation. While the results are not yet available, the study aims to be able to provide concrete, practical information and tools including model policies, strategies, and standards that will ensure the authenticity of electronic records.

Ellis Weinberger's presentation addressed what is often believed to be one of the more difficult, non-technical aspects of digital preservation: intellectual property rights. In his capacity as Cedars Project Officer, Weinberger works with Catherine Seville, Intellectual Property Adviser to the CEDARS project, to help develop collection management policies, including intellectual property rights policies for the preservation of digital materials. His presentation provided a review of rights and issues, which may arise in digital preservation activities from the more generally understood concept of general rights and legal deposit, to newer concepts such as 'database rights' and the European Union Copyright Directive. Based on his experience with the Cedars Project, Weinberger offered suggestions, advice, and even a draft to assist in rights negotiation for Digital Preservation.

Authenticity and Authentication for Digital Preservation

One of the key components of preservation is the ability to authenticate a document or perhaps, the user of a document. In the analog world, the authentication procedures are well-established and easily tracked, but in the digital world, authentication and authenticity can prove to be very difficult issues. Presentations in this session were made by Nancy Brodie, National Library of Canada; Kevin Ashley, University of London Computing Centre, and George Barnum, US Government Printing Office.

Nancy Brodie addressed issues related to *Authenticity, Preservation and Access in Digital Collections*. According to Brodie, recent studies by the National Library of Canada with the Government of Canada, the Canadian Association of Law Libraries, and the Humanities and Social Sciences Federation of Canada related to the authenticity of digital information, have shown that preservation is closely associated with authentication in the minds of scholars. Similarly, the Government of Canada (GOC) recently enacted legislation to support electronic commerce, however, the electronic publications portion of the legislation can only be brought into force "when the appropriate technology is in place for ensuring the integrity of the electronic versions." Issues related to the chain of authenticity in digital

materials were the crux of Brodie's presentation.

Kevin Ashley's presentation, "I'm Me and You're You But Is That That?" focused on security and authenticity of digital resources and the ability to draw parallels to the world of traditionally preserved materials. To Ashley, issues of security and authenticity are closely related. As 'service providers,' the identity of those who access material could be a security or an access/authenticity problem. As well, users are concerned with the authenticity of the objects they are viewing. But in a world where distributed resources are served up over networks, issues of security and access become crucial but difficult links in the chain of archiving documents. Using the title of his talk to illustrate that preservation is '"always tied up with access," Ashley described approaches and mechanisms that have been developed to provide service providers and users with appropriate levels of security and access.

George Barnum's presentation discussed the creation of the new Federal Depository Library Program Electronic Collection. Based on the legal mandate to provide free, unrestricted public access to the publications of the US Government, the Electronic Collection is refining the 'old model' of the US Government Printing Office (GPO) to create a comprehensive, virtual library of government information. With this change, however, a large challenge looms as the government must create an operational structure around the basic policy framework of assuring official integrity of information and also keeping it available and accessible permanently. To meet those needs, the government will likely need to partner with institutions now a part of the traditional depository program, such as universities, as well as with vendors and public agencies who can commit to the permanence of their publication on the original site. Other challenges related to in-house systems remain, but an effort by the GPO is underway to realize the full potential of the digital library concept for US Government information.

Making Strides: Working Together

The value of collaboration in digital preservation initiatives—inter-institutionally or internationally—was a theme echoed throughout the two days of the conference, so it was only fitting that the final session specifically addressed this. Robin Dale, the Research Libraries Group, and Neil Beagrie, JISC Digital Preservation Focus, highlighted two different collaborative efforts.

Robin Dale outlined the scope of work being covered through *International Collaboration in Digital Preservation Metadata*. Beginning with a brief discussion of the OCLC/RLG Working Group on Metadata for Digital Preservation, Dale highlighted international efforts toward establishing metadata sets for both the high-level framework (OCLC/ RLG, Cedars, NEDLIB, etc.) and granular (NISO, Audio Engineering Society, etc.) levels. Next, she reviewed the projects and institutional efforts covered during the one-day Information Infrastructures Workshop that preceded the Preservation 2000 conference. Finally, Dale summarized the preservation metadata issues that remain, including addressing familiar questions of "How much metadata is enough?" and "How much will it cost?" The issues, identified by presenters and workshop participants, will be followed up by projects represented at the workshop.

Neil Beagrie is the recently appointed Assistant Director of Preservation for the JISC Distributed National Electronic Resource (DNER) and the JISC Digital Preservation Focus. The DNER is a managed digital collection and environment for accessing quality-assured information resources on the Internet and its main objective is to stimulate the use of a collection of high quality digital resources within all areas of the UK higher and further education community. Beagrie outlined the program and activities of the JISC Digital Preservation Focus, those of the proposed Digital Preservation Coalition, and the rationale that has led to these developments to foster collaboration.

Closing Keynote Address

Jim Michalko, President, Research Libraries Group, delivered the closing keynote address. After sharing his bids for some of the more interesting phrases and salient points of the conference, Michalko bade participants to think about preservation and long term accessibility of digital materials in terms of three questions:

What have we done so far?

Why is this so hard? and

Where should our efforts go next?

While difficult to quantify what has actually been "done," (as well as what "done" means) the conference was an important step in sharing and disseminating information about key projects and research underway. With this knowledge, Michalko asserted, organizations and institutions can move to tackle important issues which remain unaddressed or unsolved. Outlining a series of necessary tools and guidelines he considers key to developing organizational digital archiving capacity, he said he was encouraged to hear several are underway in projects represented at the conference. Still, Michalko believes efforts should begin to focus on archive attributes, selection issues, organizational models and business models.

Finally, he highlighted several proposed actions raised in Lynne Brindley's keynote speech. Michalko endorsed the establishment of a UK Digital Preservation Coalition and urged participants to develop national digital preservation strategies in an international context and to act in concert to garner appropriate, vital funding for digital preservation.

Conference papers

Conference papers will be available on the RLG web site by December 19, 2000.

Footnotes

(1) National Library of Australia. *Preservation Metadata for Digital Collections: Exposure Draf*t, 2000. http://www. nla.gov.au/preserve/pmeta.html (back)

(2) CEDARS Project. *Draft Specification: CEDARS Preservation Metadata Elements*, 2000. http://www.leeds.ac.uk/ cedars/MD-STR~5.pdf (back)

(3) Networked European Deposit Library. *Metadata for Long Term Preservation*, 2000 http://www.kb.nl/coop/nedlib/ results/preservationmetadata.pdf (back)

Highlighted Web Site

Digital Library Federation Newsletter

The DLF (Digital Library Federation) Newsletter is a new quarterly publication from the DLF Director's office, based on progress reports submitted by DLF member institutions. Although designed primarily as a means for DLF libraries to report to each other on digital library activities, much of the information provided should have a broader appeal. As major trendsetters and innovators in the digital library world, DLF members provide a summary of important initiatives in new digital library services, collections, and projects in their reports.

In addition, DLF member reports provide the data for three new registries that should form important resources for the entire digital library community. One is a calendar of recent and forthcoming digital library events hosted by the DLF or its members. Then there are two searchable databases, hosted at the University of Michigan on behalf of DLF:

1) Documenting the Digital Library is a web-searchable database of policies, strategies, working papers, standards and other application guidelines, and technical documentation developed by DLF members. Topics covered include digital preservation, digital reformatting, and metadata. The document set can be searched (free-text or fielded) or browsed by institution or subject. The collection already includes nearly 150 documents.

2) Registry of Public Access Collections is a Web-searchable database of DLF members' public-domain digital collections. See the FAQ in this issue (below) for more details about this database.

FAQ

Is there a good, comprehensive catalog of Web-accessible digitized collections available on the Internet?

This is indeed a frequently asked question. From the earliest period of library and archive digitization initiatives, when nearly every effort was experimental and had the word "project" tacked on to its name, there has been an interest in creating a consolidated list of such activity. The first digital library efforts, which predated the World Wide Web, tended to be closed systems available only on a single campus. Nevertheless, lists of projects served to inform the larger community about who was doing what, which publishers were involved, what software was being developed, and what digitization techniques were being used.

It wasn't until the mid-1990s, with the confluence of lower-cost, high-bandwidth networks, rapidly declining mass storage prices, and the development of the Web, that a large number of collections of digitized material which could be both legally and practically shared started to appear. Libraries, many sporting newly developed "gateways" to electronic resources, were hungry for entries to fill their catalogs of network-accessible resources.

In 1995, the Council on Library Resources and the Commission on Preservation and Access (CPA) conducted a joint survey of existing, in-progress, and planned library digitization activity. The results were published by CPA in February 1996 in a report by Patricia A. McClung entitled *Digital Collections Inventory Report*. Then CPA director Deanna Marcum made the following observation in the report's foreword:

> *Each time librarians, or scholars, or technology experts meet—separately or together to discuss digital projects, someone invariably raises the question, "Are there any sorts of inventories of scanning projects?" There are, it seems, several incomplete lists or promises to start such lists, but no one could point to one authoritative source.*

In the "list of lists" section of her report, McClung found some good subject-specific Web-based resource guides. She found some sites providing lists of electronic journals and other digitized texts. But other than the Clearinghouse of Image Databases, housed at the University of Arizona, there was little in the way of attempts at comprehensive listings of digitized collections.

It is now nearly five years since that report was published. The Web has seen exponential growth, and the amount of digitization activity at libraries, archives and museums has increased many-fold. Search engines have proliferated and specialized listings of Web resources on even the most obscure subjects abound. Have efforts to catalog the growing list of digitized library collections kept up?

The situation is certainly better than it was five years ago. Many institutions now have umbrella departments to handle and coordinate their own digitization efforts, and have created a catalog of local digitized collections, or at least a single Web page that consolidates access to local collections.

At a higher organizational level, some state, provincial, and national agencies have created listings of digitized collections that fall within their domain. Access information for some of these appears in the resources section, below.

Finally, there is a group of both private and institutional efforts to generate more comprehensive listings. However, none has taken on the rather daunting task of filling the role of being the "one authoritative source" that Deanna Marcum was seeking. Existing efforts face a number of serious obstacles:

- Lack of definitional clarity in the criteria for selection
  - "digital images" vs. "electronic text"
  - "collections" vs. "exhibitions"
  - "converted or digitized" vs. "born digital"
  - free, open access or public vs. pay, subscription or closed access
  - permanent vs. ephemeral
  - derived from library, archive or museum materials vs. commercial or other source
- Inadequate means for locating new sites and keeping existing entries up-to-date
- Lack of funding or institutional support
- Lack of cataloging or metadata standards
- Technical difficulty in establishing browsing and searching functions
- Lack of coordination with efforts being conducted by others
- Uncertainty about who the audience is and what functionality it needs or desires

Some of these points merit further discussion.

Before one goes about creating a catalog of digitized collections, it's helpful to know who will be using it, and for what purpose. Though digitization is now a widespread practice in libraries and archives, and information about policies and practices is more readily available, there is still a strong demand for sharing of details amongst initiatives. Those planning new digitization efforts would like to see what others working with similar materials have done. What kind of capture device was used? What file format and compression technique were chosen? What software handles searching or creates the Web interface? What metadata schemes were employed?

Access to the digitized collections themselves is the second major source of interest in comprehensive catalogs of digitized materials. In some cases, the desire is for collection level access (essentially a home page link), often as part of collection development efforts specifically geared to Web-accessible resources. Another motivation is interest in creating complementary collections or avoiding duplication. There is also increasing interest in item-level access to digitized collections.

Whether a catalog of digitized material is able to fill one or more of these roles depends heavily on the structure of the records it maintains. Unfortunately, most of the existing catalogs rely on "home brew" records, and do not adhere to any known metadata standards. Some attempt to incorporate technical details of digitization, while others include only basic content description and location data.

Not surprisingly, item level access is a rarity. Many individual sites don't provide it, and those that do often use unique and proprietary techniques that resist incorporation into a more global catalog. Until metadata standards for item level access are agreed upon and widely promulgated, efforts to provide multi-collection access to individual digitized items will be fairly crude and non-standard. The Berkeley Digital Library SunSITE Image Finder consolidates search windows for 14 different image collections on a single Web page, but doesn't facilitate cross-collection searching. More ambitious is Artcyclopedia, where a merged index of the contents of over 700 sites containing fine art images can be searched by artist name or title of work (though one can only speculate on the difficulty in maintaining such an index in the current Web environment).

The other major shortcoming of the existing digitized collection catalogs also stems from a more generalized Web failure—the absence of a widely used permanent address scheme. Rapidly shifting URLs and regular site reorganization means that even the most actively maintained catalog suffers from a certain percentage of dead links and/or outdated descriptions.

It is well beyond the scope of this FAQ to discuss the various initiatives that will someday make a truly comprehensive, accurate, global digital library catalog a reality. In the meantime, here is a representative survey of some currently available tools that attempt to catalog different slices of the digitized world, despite the difficulties presented by the not yet mature Web environment. Even taking the differences in listing criteria into account, none of these sites can be considered comprehensive. There is, of course, some overlap amongst the listings within these sites, but there is also something approaching mutual exclusion between the sites focusing on collections and those focusing on exhibitions.

Finally, though there is some evidence that the maintainers of these sites are aware of each other's existence, there is, as yet, little apparent cooperation or sharing of effort.

Principal Digital Collection Inventory Web Sites

Site Name	Site Info	Date Established & Institution	Inclusion Criteria	Approx. # of Listings	Type of Data Provided	Search/ Browse Functions	Update Schedule	Notes
Clearinghouse of Image Databases	Author's Brief	1994 University of Arizona Library	Broad. See Content and Scope in Author's Brief	150	Most include contacts, description, etc. Some include technical data	Browse by institution or media, search by keyword	Relies on submitters to update their own entries	1
Digital Initiatives Database	Description	1998 Assoc. of Research Libraries & U. of Illinois at Chicago	Wide variety of digital projects involving libraries	400	Contacts, description, subject keywords, URL	Browse by project or institution, search by multiple criteria	Relies on submitters to update their own entries	2
Digital Collections Registry	DLF Registries	2000 Digital Library Federation & U. of Michigan	DLF member collections of converted library materials	300	Institution, description, source, URL	Browse all or by institution, simple search and advanced search	Quarterly	3
Directory of Digitized Collections	About Directory of Digitized Collections	1999 IFLA/ UNESCO & The British Library	Converted cultural heritage collections	150	Institution, brief description, media, language, URL	Browse all simple search advanced search	Ongoing	4
Library and Archival Exhibitions on the Web	Introduction	1995/1998 Smithsonian Institution Libraries	Online exhibitions from libraries & cultural institutions	1200	Institution, URL	Browse by title	10-12 times per year	5

Other Useful Digitized Collection Metasites							
Site Name	Institution	Site Info	Inclusion Criteria	Approx. # of Listings	Type of Data Provided	Search/ Browse Functions	Notes

Australian Digitisation Projects	National Library of Australia	Description	Converted collections, excluding exhibitions, see scope	90	Contacts, content, description, timetable, budget, URL	Browse, free text search	
Inventory of Canadian Digital Initiatives	National Library of Canada	Description	Converted collections and new resources, excluding current publishing, see scope	50	Contacts, description, genre, subject keywords, URL	Query, complex search	
Funded Projects in the DFG Programme	Deutsche Forschungs-gemein-schaft (DFG) Germany	Inventory of primary Web resources	Converted collections from research libraries	40	Contacts, timetable, short description, URL	Browse	
Colorado Digitization Project	Colorado Digitization Project	Description	Digitization projects in the state of Colorado	30	Institution, description, subject keywords, URL	Browse by format type, browse by location	6
Existing Digital Image Archives	TASI (Technical Advisory Service for Images) UK		Broad. Some focus on UK-based collections	80	Title, description, incl. some technical details, URL	Browse by broad subject category	7
Digital Librarian: Images	Private		Mostly converted collections from libraries & other cultural institutions	330	Title, description, URL	Browse	8
SunSITE: Other Digital Image Collections	U. of California at Berkeley Digital Library SunSITE			50	Title, brief description, URL	Browse	9

Footnotes

(1) Completeness of entries varies considerably. Many entries are from 1994 and lack URLs. (back)

(2) Not restricted to ARL members or research libraries generally. Seen as a planning and information sharing tool for ARL and non-ARL libraries. (back)

(3) Digital Library Federation members are required to submit twice yearly updates. DLF is interested in studying ways to improve item level access across collections and supports the Open Archives Initiative. (back)

(4) Just emerging from the test phase. The database is being moved from a British Library server to UNESCO after which it will be publicized much more widely and an active web trawling process will seek out existing collections to add. Within the boundaries of its inclusion criteria, UNESCO hopes this database "will provide a single focal point of information on digitized collections." (back)

(5) Originated at the University of Houston Library in 1995 and later taken over by Smithsonian Libraries in 1998. May add search functionality in the future. The difference between an exhibition and a collection remains somewhat indistinct, but Smithsonian Library's listing concentrates on sites which offer digitized images accompanied by descriptive narrative and identifying information on a particular theme. (back)

(6) The Colorado Digitization Project also maintains a useful listing of other statewide and national digitization efforts in the United States. (back)

(7) See An Image Collections Registry for UK HE for details about the underlying structure of TASI's database. (back)

(8) Also includes links to image resources such as directories, search engines, museums, archives, searchable databases and sites which provide information about the process of digital imaging. Additional digital collections are listed in the Electronic Texts portion of the site. (back)

(9) See also Other Digital Text Collections. (back)

—RE

Calendar of Events

Joint Conference on Digital Libraries
Call for Full Papers: Due January 9, 2001
Sponsored by the Association for Computing Machinery and the Institute for Electrical and Electronics Engineers, this first joint conference on June 24-28, 2001 in Roanoke, VA, will be an international forum focusing on digital libraries and associated technical, practical, and social issues. Participants will be from the full range of disciplines and professions involved in digital library research and practice.

The Open Archives Initiative Meeting
January 23, 2000, Washington, DC
This initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The goal of this architecture is to provide an easy way for data providers to share their metadata, and for service providers to access that metadata, and use it as input to value-added services. A key component of the

interoperability architecture is the use of the Dublin Core element set as the required resource discovery vocabulary.

American Association for History and Computing Annual Meeting:
Moving Clio into the New Millennium: Interaction, Visualization, Digitization, and Collaboration
February 1 - 3, 2000, Indianapolis, IN
The annual conference will focus on historical documents in the digital age, creating digital archives, and imaging history collections.

Digitization for Cultural Heritage Professionals
March 4 - 9, 2001, Houston, TX
The Humanities Advanced Technology and Information Institute (HATII) and the Fondren Library at Rice University are offering a one-week intensive program. The course introduces skills, principles, and best practices in the digitization of primary textual and images resources, with strong emphasis on interactive seminars and practical exercises.

Moving Theory into Practice:
Cornell University Announces Digital Imaging Workshop Series for 200l:
Registration now open for May 14-18 workshop
Offered by the Cornell University Library, Department of Preservation and Conservation, this workshop series aims to promote critical thinking in the technical realm of digital imaging projects and programs. This week long workshop will be held three times in 2001 (May 14-18, July 23-27, and October 1-5). Each session is limited to 16 individuals. Registration is now open only for the May session. The registration dates for the July and October workshops are noted on the workshop Web site. The National Endowment for the Humanities funds the workshop.

Announcements

Visual Arts Research, Learning and Teaching Collections
The Visual Arts Data Service is now offering system enhancements and new collections from its online catalogue. These include free-text selectable cross-searching of image database collections and displaying thumbnails that enlarge to screen-size images with accompanying catalogue records.

New Listserv on Digital Preservation
This listserv will carry announcements and information on activities relevant to the preservation and management of digital materials in the United Kingdom. It will be used to disseminate information on the work of the Joint Information Systems Committee (JISC), Digital Preservation Focus, and the Digital Preservation Coalition.

Hotlinks Included in This Issue

Feature Article 1
Barco Personal Calibrator(r) monitor http://www.barco.com/display_systems/products/mrkcis.htm
Collections Conservation Laboratory http://preserve.harvard.edu/collectionsconservation/labtour/start.html
Fujitsu M3097 D/G scanner http://www.fcpa.com/product/scn/scn_m3097dedg_info.html
GTI Soft Proofer(r) viewing booth http://www.gtilite.com/

Harvard and Radcliffe http://hul.harvard.edu/ldi/html/harvard_radcliffe.html

"Harvard's Library Digital Initiative: Building a First Generation Digital Library Infrastructure,"http://www.dlib.org/dlib/november00/flecker/11flecker.html

Harvard Web Site http://preserve.harvard.edu/dig/

Hedda Morrison photographs of China http://hul.harvard.edu/ldi/html/harvard-yenching.html

HUL Library Notes (PDF) http://hul.harvard.edu/publications/library_notes/pdfs/HULN1289.pdf

Informatik, Inc. http://www.informatik.com

LDI Internal Challenge Grant Program http://hul.harvard.edu/ldi/html/grants.html

Library Digital Initiative http://www.fcpa.com/product/scn/scn_m3097dedg_info.html

Library Preservation at Harvard "Image Digitization" http://preserve.harvard.edu/resources/digitization/image.html

Library Preservation at Harvard Web Site http://preserve.harvard.edu/standards/imagequality.html#viewing

Nuremberg Military Tribunal http://www.law.harvard.edu/library/digital/war_crime_trials_nuremberg.htm

Scitex Leaf Volare digital camera back http://www.creoscitex.com/products/input/leaf/volare/volare.asp

Feature Article 2

5th Framework Programme http://www.cordis.lu/fp5/home.html

Arts and Humanities Data Service http://ahds.ac.uk

Bibliotheque Nationale de France http://www.bnf.fr

The British Library http://www.bl.uk

CAMILEON Project http://www.si.umich.edu/CAMILEON/

CARMEN AP 2/5 http://harvest.sub.uni-goettingen.de/carmen

CARMEN Project http://www.mathematik.uni-osnabrueck.de/projects/carmen

Cedars Project http://www.leeds.ac.uk/cedars

Department of Preservation and Conservation http://www.library.cornell.edu/preservation

Digital Object Identifier http://www.doi.org/

Dublin Core 1.1 http://www.dublincore.org/

<indecs> http://www.indecs.org/

InterPARES Project http://www.interpares.org

LOCKSS http://lockss.stanford.edu

METAe http://meta-e.uibk.ac.at/

National Library of Australia http://www.nla.gov.au

National Library of Medicine http://www.nlm.nih.gov

Networked European Deposit Library (NEDLIB) Project http://www.kb.nl/kb/ict/dea/index-en.html

PANDORA Archive http://www.nla.gov.au/pandora

project DNEP http://www.kb.nl/dea/index-en.html

Project PRISM http://www.prism.cornell.edu/

OCLC http://www.oclc.org

Open Archival Information System http://ssdoo.gsfc.nasa.gov/nost/isoas/us/

RLG Web site http://www.rlg.org/toc.html

Highlighted Web Site

Digital Library Federation Newsletter http://www.clir.org/diglib/pubs/newsletter.htm

DLF member institutions http://www.clir.org/diglib/about.htm

Documenting the Digital Library http://www.hti.umich.edu/cgi/d/dlf/dlf-idx

Forthcoming digital library events calendar http://www.clir.org/diglib/pubs/events.htm
Registry of Public Access Collections http://www.hti.umich.edu/cgi/d/dlfcoll/dlfcoll-idx

FAQ

Artcyclopedia http://www.artcyclopedia.com/index.html
Australian Digitisation Projects http://www.nla.gov.au/libraries/digitisation/projects.html
Clearinghouse of Image Databases http://www.library.arizona.edu/images/clearinghouse/clearinghouse.html
Colorado Digitization Project http://coloradodigital.coalliance.org/colorado.html
Digital Collections Inventory Report http://www.clir.org/cpa/reports/mcclung/mccllong.html
Digital Collections Registry http://www.hti.umich.edu/cgi/d/dlfcoll/dlfcoll-idx
Digital Initiatives Database http://db.arl.org/did/review.html
Digital Librarian: Electronic Texts http://www.digital-librarian.com/electronic.html
Digital Librarian: Images http://www.digital-librarian.com/images.html
Directory of Digitized Collections http://thoth.bl.uk/ddc/index.html
Existing Digital Image Archives http://www.tasi.ac.uk/resources/img-resources.html
Funded Projects in the Deutsche Forschungsgemeinschaft (DFG) Programme http://www.sub.uni-goettingen.de/gdz/en/projects/vdf.en.html
An Image Collections Registry for UK HE http://www.tasi.ac.uk/2000/10/ICR/project/ICR_project.htm
Inventory of Canadian Digital Initiatives http://www.nlc-bnc.ca/initiatives/erella.htm
Library and Archival Exhibitions on the Web http://www.sil.si.edu/SILPublications/Online-Exhibitions/online-exhibitions-frames.htm
Open Archives Initiative http://www.openarchives.org/
State and National digitization efforts http://coloradodigital.coalliance.org/other.html
SunSITE: Other Digital Image Collections http://sunsite.lib.berkeley.edu/Collections/otherimage.html
SunSITE: Other Digital Text Collections http://sunsite.lib.berkeley.edu/Collections/othertext.html
SunSITE Image Finder http://sunsite.berkeley.edu/ImageFinder/

Calendar of Events

American Association for History and Computing Annual Meeting http://www.theaahc.org/annual.htm
Digitization for Cultural Heritage Professionals http://www.rice.edu/Fondren/DCHP01/
Joint Conference on Digital Libraries http://www.jcdl.org
Moving Theory into Practice http://www.library.cornell.edu/preservation/workshop/
The Open Archives Initiative Meeting http://www.openarchives.org/DC2001/OpenMeeting.html

Announcements

New Listserv on Digital Preservation http://www.jiscmail.ac.uk/lists/digital-preservation.html
Visual Arts Research, Learning and Teaching Collections http://vads.ahds.ac.uk/search.html

---

Publishing Information

internationally via the RLG 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 jlh@notes.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, Carla DeMello.

All links in this issue were confirmed accurate as of December 13, 2000.

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

CONTENTS   SEARCH   HOME

Trademarks, Copyright, & Permissions