CASE STUDIES

The following case studies represent only a small fraction of the library projects identified, but they are among the most interesting.

Pioneering Efforts

The first group of case studies are included because these organizations were among the pioneers five and more years ago.

Columbia University Avery Library

One of the earliest imaging projects undertaken by a library was Project AVIADOR (Avery Videodisc Index of Architectural Drawings on RLIN), launched by the Avery Library of Columbia University in 1985. The primary goal of the project was to provide machine-readable cataloging information on 41,000 architectural drawings and incorporate an image of each drawing on an electronic storage device. The intent was to allow users to flip through whole collections or go directly to a specific drawing, all without handling the originals. Not only would access to Columbia's historic collections of architectural drawings be significantly improved, but the project would help to preserve them.

The drawings in the initial project represent 40 prominent archival collections at Avery, including the works of such well known architects as Frank Lloyd Wright (c.546 items), those of Wright's mentor, Louis Henry Sullivan (c.175 items), those of the popular renderer, Hugh Ferris (c.310 items), and a sample of McKim, Mead & White (c.560 items). The smallest collection cataloged during the project is that of James Kenwick, Jr. (58 items) and the 8,248 drawings from the Ely Jacques Kahn collection represent the evolution of his firm over many years.

AVIADOR was launched when imaging technology was still new. Rather than relying on relatively new digital scanners and optical storage devices, the decision was made to photograph the materials and store them on analog videodisc with the intent to create digital images from the high-quality photo or negatives at a later date.

AVIADOR was designed for use with a personal computer connected to the Research Libraries Information Network (RLIN) mainframe. RLIN devised a program for the interface that connected each of the 41,000 still-frame images of drawings on the videodisc to the appropriate catalog record on RLIN.

The user could:

1. Search for the appropriate bibliographic record(s) in RLIN using a variety of textual indexes (such as architect, title, building name, geographic location, and so on.)
2. When a desired record is identified and displayed in the IBM PC monitor, place the cursor at the accession number(s) for a particular drawing within the RLIN record
3. Then call up the corresponding visual image(s) on the videodisc monitor by pressing the appropriate function key on the IBM PC keyboard.

Alternatively, the user could:

1. Begin with the graphic index by scanning the images on the videodisc
2. When a desired image has been identified and displayed on the video monitor,

retrieve the corresponding bibliographic record by pressing the appropriate function key on the IBM PC keyboard.

Fifty copies of the videodisc were produced at $500 each, including the necessary software and a user manual.

The image files were subsequently converted to digital format and stored on magnetic disk drives.

Ford Motor Technical Information Center Library

One of the first imaging files linked to an automated library system was that of the Technical Information Center (TIC) Library, the major scientific and engineering library for the Ford Motor Co. In 1989 it sent out an RFP for a document-imaging system that would provide users with desktop electronic access to full-text of company technical reports and would also be integrated with the Library's online catalog. The Library was already using COMSTOW Information Services' BiblioTech online integrated library software for all library functions.

The TIC staff had long recognized a need within Ford Motor Co. for a central depository of company reports for many users. The Library had long maintained a collection of Research and Product & Manufacturing Engineering staff reports, but there was no central location for reports generated by other research or engineering groups. As a result, valuable information was not shared as widely as it should be. The TIC staff saw optical storage and imaging technology as a means of providing the entire organization with access to company technical information without increasing the Library's need for physical space or adding another place to look for information.

The BiblioTech system was mounted on a Digital Equipment Corp. (DEC) MicroVax 3900 with 32 Mb of RAM and more than 2 Gb of magnetic storage. The successful bidder proposed capturing the images using digital scanners and storing them on CD-ROM. The Library accepted the proposal and subsequently implemented an interface between the BiblioTech system and PC-based CD-ROM devices containing the digital images.

Given their experience with access to CD-ROM, the TIC staff concluded that any imaging system should build on, and be totally integrated with, existing information management systems. In Ford's case, this meant that an imaging system must serve as back-end system to the BiblioTech system and BiblioTech would serve as the indexing-finding tool. When a document is located in BiblioTech, the availability of a full-text image would be indicated in the record, and a simple keystroke sequence would clear the screen and bring the image to the user. The user could then move through the document, panning and zooming as necessary. After viewing a document, users could return to BiblioTech at the exact point where they were before viewing an image. Users would be able to print imaged documents at their workstation or fax them to their desktop. Several years passed before the interface became seamless.

National Agricultural Library Text Digitizing Program

A microcomputer-based scanning system was installed at NAL in January 1988. More than 4,000 pages of noncopyrighted aquaculture publications were scanned and digitized to create both bit-mapped page images and ASCII text. The text was indexed using TextWire Plus from Unibase, and the resulting database, "Aquaculture I", was distributed on CD-ROM to the participating Land Grant Libraries in March 1989. The libraries' role was to evaluate the delivery medium, the retrieval software, and the contents of the disk itself. In addition, a comparison of retrieval from ASCII text alone versus ASCII text with page images was conducted.

A second disk, "Food, Agriculture and Science," was produced by the Consultative Group on International Agricultural Research (CGIAR) and sent in September 1989 to both the Land Grant sites and to CGIAR sites around the world for evaluation. It contained CGIAR materials nominated by the CGIAR sites themselves, uses KAware2 retrieval software from Knowledge Access, and includes both full text and graphics. The project was so successful that the CGIAR commissioned the production of an entire series of CD-ROMs.

In May 1990, a third database was distributed: A two-disk set produced by the University of Vermont, consisting of Canadian Acid Rain publications. Two Title II-C grants from the U.S. Department of Education to identify, scan, and catalog the publications supported this effort. The materials, together with copyright permission, were made available by various agencies of the Canadian government. These CD-ROMs, entitled "Acid Rain: Canadian Government Documents," use the KAware2 retrieval software and contain full-text and full-page images.

The fourth and final evaluation disk contained more than 4,100 selected pages from NAL's large special collection on Agent Orange. Produced by NATDP and entitled simply "Agent Orange," it uses Windows Personal Librarian retrieval software from Personal Library Software, and contained the full text of the selected documents, 4,100 high-resolution page images, and a sample of 200 low-resolution page images. It was distributed in October 1990.

A disk entitled "Food Irradiation" and running under TextWare Plus retrieval software, was mastered in June 1991 and used enhanced bibliographic records for search and retrieval. Like the full-text documents contained on the evaluation disks, the bibliographic records had links to their associated page images. This was NATDP's first experiment with the use of bibliographic records for retrieval. The 5,100 pages on this disk were from a large collection housed at NAL, and the source publications had been donated to a third party, freeing up much-needed space in the library.

The NATDP Evaluation Study Final Report recommended continued production of electronic publications but stipulated that future products include only image files; full-text file conversion was deemed too time-consuming and costly. NAL promptly embarked on additional electronic publications. For the new program it licensed state-of-the-art retrieval software, Windows Personal Librarian from Personal Library Software of Rockville, Md. The first disk, issued in May 1992, consisted of Volumes 1 through 16 of the *Agronomy Journal*, dated 1907 through 1924. The database was made up of bibliographic records and page images. It was produced with the cooperation of the American Society of Agronomy.

The next disk, released in late 1992, was "Aquaculture II." It contained 6,500 page images. After that came the "George Washington Carver" CD-ROM, containing 3,500 page images scanned from three reels of microfilm of Carver's papers, letters, and illustrations. The microfilm is part of a 67-reel collection produced by Tuskegee University. This is one of the first major scanning projects by a library from microform.

The NAL subsequently installed two VTLS InfoStations to facilitate the linking of its image databases with its VTLS automated library system. The InfoStations consisted of NEXT workstations with 2.4 Gb of hard disk. Images were selectively downloaded from the CD-ROM onto the NEXT hard disk storage. Users search the bibliographic database on the VTLS system by author, title, subject, or other access point at the InfoStation, determine that an image file exists for a bibliographic record, and retrieve the image onto the InfoStation screen by clicking the "Retrieve Multimedia" button at the bottom of the Windows display. Users can return to the bibliographic record screen when finished with the image file. The interface was seamless—users need not log off one system to access the other, nor do they lose their place. However, there is one drawback—you can't tell if an image exists for a bibliographic record by consulting the initial brief bibliographic record display; you must call up the full bibliographic record.

Peabody College Education Library (Vanderbilt University)

The Clipper Project of the Peabody College Education Library was a Mac-based optical storage system to allow easy retrieval of active reference files by end-users. Clipper used Macintosh computers both as file servers and workstations. MARS software from Micro Dynamics, Ltd., through an interface customized for each of its clients, managed the storage and retrieval of the documents. The installation, made in 1988, was one of a few turnkey systems installed in a library by a major multi-industry imaging vendor. The project was funded in part by a hardware grant from the Apple Academic Development Donation Program.

The images—some 12,000 of them—were scanned from materials in 10 vertical filing cabinets: newspaper clippings, newsletter articles, pamphlets, working papers, reports, and so on. Much of the information was current and local, from Tennessee, Nashville, or southern states. It also included lots of statistical information in the form of charts, graphs, tables, and statistics. There were also items of national interest related to education. Before Clipper, you could retrieve this type of information only by looking through the individual file folders by a single subject heading.

The scope for Clipper was much narrower and more focused than ERIC or PsycLIT. It did not contain comprehensive material on a subject. For example, if you were interested in the topic of corporal punishment, Clipper might contain articles that focused on Metro or Tennessee's use of corporal punishment. However, the user would still need to search other indexes or databases to research this topic completely.

The major components of Clipper were:

- Input workstation—consisting of a Macintosh IIx with an Apple scanner. Documents were scanned, their images temporarily stored on the hard disk, and descriptive information about each document entered at the workstation.
- Optical Server—a Macintosh II with a Pioneer 5.25-inch Double WORM Drive attached. The Optical Server managed the storage and retrieval of scanned images that were archived from the input workstation's hard drive to WORM disks.
- Directory Server—a dedicated Macintosh SE/30 that stored descriptive information about the scanned images stored on the WORM disks. It acted as an index to the stored images.
- Retrieve workstation—a Macintosh IIx with a 19-inch Sigma Designs monitor and LaserWriter SC for printing retrieved images. The workstation was set up in the reference room.
- MARS (Multi-user archival and retrieval system)—software that allowed the network of Macintoshes to store, find, and retrieve document images.
- An Ethernet LAN that connected the Clipper system to the campus network, making Clipper available in offices and rooms outside the library.

AT&T

Since 1989, the AT&T Information Services Network has been scanning internal technical memoranda, initially at 400 dpi (dots per inch) and storing the images on write-once-read-many (WORM) optical disks. Tens-of-thousands of documents were scanned, and new documents continue to be scanned. Initially, requests for copies were filled by printing the document on a 400-dpi printer at a central site and sending it out via company mail. Although this system was a large improvement over its predecessor—where a clerk located the original document in a filing cabinet and made a xerographic copy—there was still a several-day delay before the requested copy was received by the requestor.

When the AT&T network bandwidth was increased to 456 Mbps all constraints on the movement of images from the image server to desktop workstations was removed. Images are now retrievable via the company network.

Boulder Public Library

Boulder Public Library (BPL) implemented photo image access in its online catalog in 1992, an enhancement to its CARL System Public Access Catalog. Funded in part by the Boulder Public Library Foundation, the Photo Image Access Project offered a new type of access to visual materials. Designed originally to be used with the Historical Photograph Collection of the BPL, the process was subsequently applied to other visual collections as well.

The process linked photo images to corresponding MARC records in BPL's bibliographic database, and allowed the PAC user to display the images on demand on a PC.

Project staff at BPL selected photos to be scanned from the collection and determine the bibliographic identification (BID) number for the corresponding MARC record via CARL Systems' Bibliographic Maintenance. After determining scanning resolution according to a chart developed for the project by CARL, the staff member scanned the photo, stored it in its own unique file, and saved the file, named with the appropriate BID. The image was stored in compressed format, at 640 by 480 pixels with a minimum of 32 gray values. The process supported both color and black-and-white photos, although color was slower to scan and required greater storage capacity.

To look at the photo image, the user searched the PAC as usual. Image display required a PC workstation with CARL Systems emulation software and a VGA monitor. When a bibliographic record with an image linked to it was retrieved, the online message "imaged" appears as a part of the short display. The user could then select the full bibliographic record and, by responding to an additional prompt, display the image on the PAC screen.

BPL stored the photographic images locally on magnetic disk and made them available to PAC terminals connected to BPL's local area network.

The scanning process used a Microtek 300z scanning device and Picture Publisher software. The scanning software ran in the Microsoft Windows environment. The scanning station required a 33 MHz 80386 PC-compatible machine with 8 Mb of memory and 200 Mb of disk, and a mouse.

CARL Systems and the Boulder Public Library Foundation jointly marketed the imaging process and linking software to CARL Systems libraries and other interested parties.

Recent Efforts

The following programs have been begun in the past few years:

Chicago Public Library

CPL obtained a major grant in 1997 to capture images of its extensive collections of historical photographs, manuscripts, and maps. The goal was to put interesting collections on the Web, rather than preservation. The source documents were all black-and-white, so they were scanned at 300 dpi on an Agfa DocuScan flatbed scanner—a scanner capable of 36-bit color and 600 by 1,200 dpi. An archival TIFF copy was produced and so was a JPEG for Web-access.

The interesting aspect of CPL's digital collections is that they have been arranged thematically to tell a story. In fact, the collections are called digital exhibits to underscore they are not just images of entire collections. Among the exhibits are the great ships of the Great Lakes, Chicago's harbor, and so on.

Library of Congress

For those who would like more information, the best place to start is with a list of FAQs (frequently asked questions) at **http://memory.loc.gov/ammem/helpdesk/amfaq.html**.

Almost everyone is familiar with the Digital Library Program of the Library of Congress (LC) and its American Memory Historical Collections component, therefore, LC's effort will not be included as a case study.

There are many links from the FAQs pages.

Note that the LC has used service bureaus for most of its image capture. One of its RFPs is available on the Web. The URL is excessively complex, so the best way to reach the document is by searching under "Library of Congress digital images RFP."

Library of Virginia

The Library of Virginia, a state agency that provides reference and research assistance not only to state government but also to the libraries and citizens of the state, has built one of the largest image collections in the nation. There were more than 2.2 million images on its Web site as of late 2000, including court records, manuscripts, maps, and photographs. The collections of several public libraries are also included.

The Library of Virginia not only does its own image capture, it also uses several service bureaus. It prefers to capture images at 300 to 400 dpi, but occasionally go to 600 dpi for maps and manuscripts. Few of the materials are in color so shooting at a higher resolution than 600 dpi isn't perceived as necessary. It uses a Fujitsu ScanPartner 600 flatbed scanner, a Fujitsu 9096 Plus, and a Zeutschel 5000—the last a machine with a cradle to protect bound books. The first captures images in PDF format for direct loading onto the Web; the other two capture images in TIFF format. Although the image server was supplied by VTLS, the linkage from the MARC-formatted bibliographic records on the VTLS automated library system is through the 856 tag, therefore, either the image server or the bibliographic database server can be replaced without affecting the other.

University of Pennsylvania

The Schoenberg center for Electronic Text and Image is one of the most ambitious of the academic library imaging programs. The center was established to provide the scholarly community with Web access to facsimiles of original texts, documents, photographs, maps and other materials from Penn's collections. Much of the image capture has been done in-house using a variety of equipment, including three HP 4C scanners, an HP ScanJet ADJ scanner, a Polaroid SprintScan 35 side scanner, Kontron ProgRes 3012 planetary scanner, and a Phase One StudioKit camera. Much of the effort of the staff of three professionals and five student project technicians over the next two years will be focused on "Shakespeare and the English Renaissance," a planned major archive of virtual Shakespeareana; and a contract with UMI to capture the text of 200 titles selected from Pollard & Redgrave and Wing. These scans will eventually be incorporated by UMI into a product based on digitizing one of their larger microfilm series. Lesser ongoing projects include the continued capture of corporate annual reports and American salesmen's sample books from the 19th and early 20th centuries. Given the wide range of source documents, capture is done at 300 to 600 dpi and sometimes higher.