Applications of digital optical disks in library preservation and reference

by WILLIAM R. NUGENT
Library of Congress
Washington, D.C.

ABSTRACT

A Library of Congress pilot project is described in which digital optical disks will be used to provide high-density storage of textual page images. The configuration of the system is shown and the advantages of optical storage for preservation of library materials is explained. It is further shown that while library preservation is the primary objective of the project, the characteristics of digital optical disk provide simultaneous advantages in online reference retrieval. The project plan is described; it calls for scanning and storing one million pages of text over a two-year period. The new system is compared to the library’s existing digital optical disk system for demand printing of catalog cards. The future of digital optical disks in libraries is then discussed.
INTRODUCTION

The deterioration of library materials has by now brought about a crisis in all large research libraries. A significant fraction of such libraries' collections of print on paper, over 25 percent by some estimates, are exhibiting the symptoms of terminal acidic decay: turning brown, becoming brittle, and ultimately crumbling into flakes and dust. The cause is well known: the increase in demand for printed matter in the last half of the nineteenth century caused paper manufacturers to make paper from wood pulp and sizing materials that together, by acidic hydrolysis, ultimately form sulfuric acid, which in turn breaks up the normally long and flexible cellulose polymers of the paper. To combat this problem, the Library of Congress has initiated two programs.

The first program, which is proving very successful, is that of gaseous deacidification of large groups of books placed in large vacuum chambers. By this means, books that are in good condition can have their life expectancies extended another 400 to 600 years. A cooperative program with NASA is currently showing the way to achieve deacidification in 5,000-book lots. The second program, the one we address here, will use the high-density digital optical storage of page images as an alternative means of library preservation. A development contract has been awarded to Teknekron Controls Incorporated of Berkeley, California, for a system that will scan and digitize text pages, store them on digital optical disks, and under access control of the library's central computer system, will retrieve desired page images for high-resolution display and printout. The digital disk medium itself has a life expectancy of only 10 to 20 years, which is a mere instant in preservationists' thinking; but by the well-established techniques of error detection and correction, early signs of increased soft errors can be detected in time to make a perfect copy on a fresh disk of the same technology or on a future digital medium of a newer technology. By this process, the page images may thus be preserved indefinitely, or for as long as they are believed to be useful.

A second problem of large libraries, of somewhat lesser severity, is that of document delivery to the library patron. With contemporary online computer retrieval systems, such as the Library of Congress' LOCIS system, patrons can find citations of desired items and their call numbers almost instantly. But the time from a patron's request for a work until it is manually searched for in the stacks and delivered can too often exceed an hour. The online accessibility of page-image material on digital optical disks can reduce this document-delivery time to seconds for materials available on optical disks. Greater integrity of the collection is also achieved, avoiding the too common "not-on-shelf" problem.

In the classical library environment, the dual objectives of preservation of materials and providing of frequent public access to them are opposed to each other. Preservation generally means a strictly controlled physical environment, watchful custodial care, and limited public usage. High public usage generally means accelerated wear and deterioration. But page images preserved on digital optical disk can now meet both objectives without conflict, since no wear results from the low-power laser beam reflecting data from the disks.

DENSITY CHARACTERISTICS OF DIGITAL OPTICAL DISK AND PAGE-STORAGE CAPACITY

The primary attribute of digital optical disks is extremely high recording density: Bit areas are generally of the same order of magnitude as the wavelength of the laser light used to record them; that is, in the order of 1 micron or less. Typical numbers for a single-sided 12-inch disk and single-sided 14-inch disk are $10^{10}$ bits and $5 \times 10^{10}$ bits, respectively. The general range is within a factor of two of these numbers. The 12-inch Thomson/CSF disk that will be used in the library's application will initially have $0.5 \times 10^{10}$ user bits, and will later be raised to $10^{10}$ user bits. We will use the latter number in our library storage examples. On a per-surface basis, the capacity of an optical disk is thus about 150 times that of a magnetic disk.

Although this results in many obvious applications of digital optical disk as a computer-room peripheral for mass storage of conventional character-encoded data, it is this observer's view that the highest potential applications are those involving digitized image storage. For the first time, a practical non-contact medium of extremely high density exists to make such image applications cost-feasible.

THE LIBRARY OF CONGRESS PILOT PROJECT

In order to establish a firm and cost-justified basis for the digital-storage approach to library preservation, the Library of Congress has established a pilot project using digital optical storage technology.

The object of the pilot project is to establish a production environment in which both capital costs and operating costs can be determined, so that decisions on expansion can be made. It was determined that scanning and storing 500,000 pages per year would constitute a project of sufficient size. Other questions to be resolved include the depth of indexing and abstracting required when full text rather than citations only are available. Of high importance as well will be life testing, both in real time and in accelerated time, of the digital
optical media and the establishment of test criteria for determining when a disk is beginning to degrade so that replication can be effected before correctable errors become noncorrectable ones. Since library preservation is the primary objective of this program, determining the best procedures to achieve this aim will be a major contribution of this pilot project.

SYSTEM CONFIGURATION

The specified configuration of the Library of Congress digital optical disk system is shown in Figure 1. The core of the system is a Video System Controller that directs all systems input/output activities and provides the functions of video-crosspoint switch and compressor/expander for the digitized images. Two optical disk units are shown, one of which will incorporate a 100-disk jukebox. Two magnetic disks serve as image buffers for materials being accessed by the display clusters and for materials being scanned by either the document-page scanner or the microfiche scanner. The Video Image Controller, a high-speed laser printer (a Xerox 5700), the scanners and a two-terminal cluster will be housed in a computer room in the library's James Madison building and will service two-terminal clusters in three library reading rooms, one in each of the library's three buildings. The initial reading-room configuration will have two user retrieval stations, each consisting of a high-resolution CRT terminal for query entry and for display of full pages of text and an associated medium-speed printer for page-image printout. The Video System Controller is also interfaced to the library's Computer Service Center and uses the library's LOCIS online information-retrieval system for locating citations to page-image material available on digital optical disk.

RELATION TO THE LIBRARY OF CONGRESS'S EXISTING OPTICAL DISK SYSTEM

The library's Cataloging Distribution Service is now using digital optical disk for storage of images of 3" x 5" catalog cards in its DEMAND system for high-speed demand printing of catalog cards. This system was developed for the library by Xerox Electro-Optical Systems and was placed in production in August 1982. An exceptionally high resolution system, it laser-scans and laser-reproduces cards at 480 lines per inch. Output is by a specially adapted Xerox 9700 laser printer that produces 12 cards per second. More than 200,000 images of master catalog cards can be stored on one side of the Xerox 14-inch digital optical disk. This is the equivalent of 140 card drawer cards. The cards output from this system are used to fulfill orders from libraries all over the world, and the DEMAND system has made this operation faster and more cost effective. The success of this first-in-production digital optical disk system has given us confidence that digital optical disk is eminently practical for the storage and retrieval of textual page-images as well as for catalog cards.

THE FUTURE OF DIGITAL OPTICAL DISKS IN LIBRARIES

We expect many future applications in libraries as digital optical disk technology is further refined and lowered in cost. In addition to benefits in preservation and reference retrieval, there are significant benefits in storage space and space is an increasing problem in all major research libraries. Today's optical disks offer a shelf-space advantage of about 200X, and we expect a 1000X advantage in a few years. As a second future development, we expect a rapid decrease in the cost of disk players that, in turn, will lead to decentralized use of these devices on a stand-alone basis for special collections, individual reading rooms, and in smaller libraries. The next improvements in greater densities will, in addition, permit the cost-effective storage and retrieval of halftone images and color plates. A particular challenge to be met in the future is that of effectively scanning and encoding pages of mixed material, such as art books that may mix black and white text, halftone graphics and color plates on one page. When this problem is solved, we will have achieved true electronic facsimiles of the printed book.

REFERENCES

5. LaBudde, Ed. “Tracking the New Storage Strategy—Optical Disk Memo-
6. Goldstein, Charles M. “Optical Disk Technology and Information.” *Sci-
August 1979, pp. 26-33.
PERSONAL COMPUTERS

Since their invention over 49 years ago, electronic computing machines have touched the lives of nearly every person on the face of the earth. This has never been more true than now, as we enter the age of personal computers. A personal computer is a dedicated resource, relating one-to-one with its user/master. Whether a production-enhancing tool at one's work, a mind-expanding diversion in the home, or a companion on the road between, a personal computer enriches the lives of those who use them.

For this year's NCC, 24 industry observers and leaders combine in eight sessions to present the 1983 picture of pertinent issues in the personal computer marketplace. Topics range from crucial industry concerns such as software transportability between processors to new-technology methods of protection against program pirates. Included also are practical how-to seminars on business planning with microcomputers and applications for the new generation of mobile carry-along processors. One of the sessions is dedicated to the computer designer's view of the microprocessor that has all but dominated new product introductions since last year's NCC—the 16/32-bit 68000 dual-processor chip.

The most quickly evolving of all computer types deserves special attention to the future, and several of this year's sessions concentrate on the personal computers of tomorrow and their continuing impact on our society. Portia Isaacson has again assembled a panel of experts to look into the future of microcomputing. Industry maker Adam Osborne and seer Jerry Pournelle will project the ways in which personal computers will continue to alter our lives during the rest of this decade.