■ LOS ALAMOS NATIONAL LABORATORY

by Lisandro R. Ramos

A Network Long Overdue

This research library takes advantage of CD-ROM networking technology to create a '"library without walls."'

Can you imagine a library without walls? Certainly, not the least of challenges that might present would be where to put the books. But, what about a paperless library? The concept might seem foreign at first, but such a configuration is far from fiction.

The Los Alamos National Laboratory (LANL), which achieved fame in the 1940s as the main site for atomic weapons building and design during World War II, is perhaps the first institution to make the paperless library a reality. LANL now operates as a center for nuclear-weapons design, spread out over a 43-square-mile campus in the remote Jemez Mountains of northern New Mexico.

LANL's mission is focused in U.S. industry and national security. As a by-product of the laboratory's research into nuclear weapons, LANL is working under the federally-mandated Technology Transfer Act to make U.S. industry more competitive in the areas of environmental controls, biomedical research, materials research, supercomputing, lasers, and superconductivity.

As a result, LANL ranks as a world leader in computing research, using the most advanced machines available, like the Cray Y-MP/832 supercomputer and the CM-5 massively parallel computer. Designated as one of two DOE High-Performance Computing Research Centers, the laboratory helps keep the United States on the leading edge of computational science.

LIBRARY CARD

The maxim, "You are only as strong as your weakest link," is undoubtedly true when applied to scientific research. Using a typical research library, a researcher might spend days or weeks searching for information using printed indexes and abstracts. With the advent of CD-ROM technology and on-line searching, research time can be substantially reduced. At LANL, researchers are finding CD-ROM formats such as bibliographic, directory, full-text, and full-image databases to be an indispensable tool for information gathering.

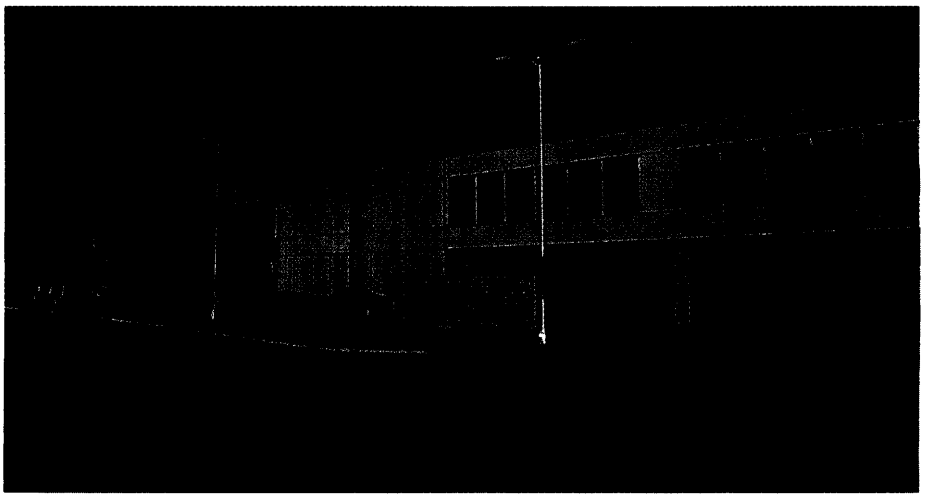
LANL's research library is a repository of scientific and technical information that comprises more than 300,000 books and bound journals, 2,000 journal subscriptions, more than 1.25 million unclassified and several hundred thousand classified reports, foreign publications, and a wide collection of CD-ROM databases supporting various disciplines.

To accommodate the growing electronic needs of the researchers at LANL, an internal network was created in *1991* and is in a constant state of expansion and refinement under the direction of manager Rick Luce, who has been instrumental in moving the library into the digital world. The internal network consists of Unix and NetWare-based servers, with CD-ROM arrays, CD-ROM and optical jukeboxes, and RAID systems. Workstations I include 80 PCs. mostlv PC/AT clones (486s and P5s), some NeXT' machines, and a few Macintoshes. All workstations are connected to an integrated 10Base2 cable plant, attached to a Cisco Systems (San Jose, CA) router for external communication. The workstations are used to access information sources both internal and external to the library, e-mail, and word processing.

FIRST LOOK

In 1990 the library began using CD-ROM technology for bibliographic search and retrieval on a standalone PC. The configuration consisted of a PC/XT and a Hitachi CD-ROM drive. The library's first CD-ROM title was an electronic encyclopedia. Later, the PC/XT was replaced with a PC/AT with five Sony CD-ROM drives used to access other CD-ROM titles.

At the same time, a network solution was being investigated for sharing CD-ROM titles with multiple discs; a decision had to be made



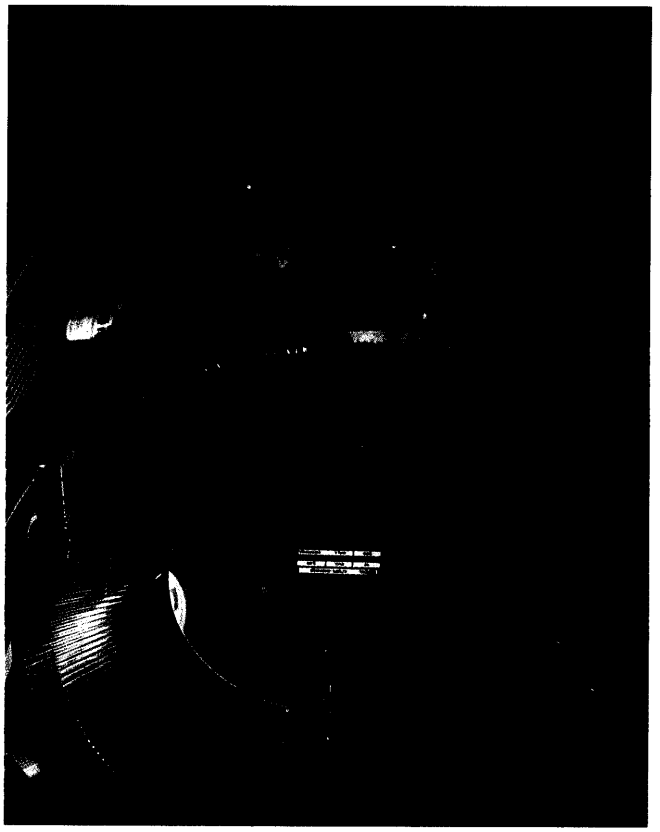
A Virtual Library. *The research library at Los Alamos National Laboratory uses CD-ROM networking technology to provide a completely digital repository of scientific information.*

whether the solution would be Unix-, Mac- or MS-DOS-based. Initially, the hope was for a Unix solution, because the laboratory's main computing machines were mostly Unix-based running over 1P. The librarians eventually decided to go with an MS-DOS solution, however, because 75 percent of all search and retrieval (or search engine) software provided by scholarly CD-ROM publishers was DOS-based, and many titles needed by the research library were only available for MS-DOS.

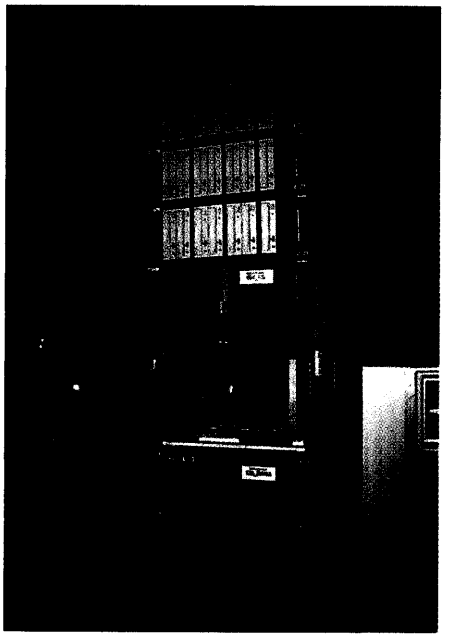
At that time, the library had only proprietary terminals through which patrons and staff accessed the on-line catalog system. In fiscal years 1991 and 1992, faced with the reality of an MS-DOS-only solution, the library staff purchased two Meridian Data (Scotts Valley, CA) CD-Net 310 turnkey systems for networking CD-ROMS; six public PC workstations were networked using NetWare 3.11. The NetWare file server consisted of a 486/33MHz machine with 6MB of RAM, and a single 320MB ESDI Seagate Technologies (Scotts Valley) hard drive for storing search engines; each CD-Net optical server consisted of a 386/20MHz machine with 8MB of RAM on a single board computer, 14 Toshiba XM-3301B SCSI CD-ROM (single-speed) drives, a monochrome video card, a 3.5-inch floppy drive and controller, and an EPROM modified Western Digital 8-bit SCSI host controller—most of which was proprietary hardware.

User and library staff PC workstations consisted of old IBM PC/XTs, ATs, and AT clones. The appropriate CD-Net drivers, based on Microsoft CD-ROM Extension and the IPX protocol, and TCP/IP software were loaded on those machines to allow access to CD-ROM databases and other external library on-line catalogs via the Internet. Access to those services was provided via Direct Access Network, a DOS-based software menuing system.

The library's first major investment in CD-ROM-based bibliographic databases was in the Institute for Scientific Information (1S1) Science Citation Index database, which is considered one of the more important databases used at the laboratory in researching science and engineering journals. The Science Citation Index database came with 12 discs, which covered issues from 1980 through 1991.



Powerful Imagery.This *image server fiom University Microfilm, Inc. (UMI, Ann Arbor, Ml) is part of a turnkey system that provides digitally scanned images of articles exactly as they appeared in the original publications. The image server has three attached CD-ROM jukeboxes, each holding up to 240 CD-ROMs.*



High and Mighty. *This 19-inch, expandable rack-mount cabinet was purchased for the existing JZe server components, which included two CPUS,* ***14*** *hard drives, and 28 CD-ROM drives.*

The popularity of the CD-ROM network grew rapidly; patrons often stood in line to access the next available public PC workstation-all of which were 486/33MHz machines with 640KB of RAM and 2MB of extended memory, each running MS-DOS 5.0 and QuarterDesk's QEMM 6.0. Consequently the CD-ROM network quickly became a high-support item for librarians.

Two common problems began to crop up. The first was the constant optimization of PC memory for loading large search engines. In order to access mounted CD-ROMs at the CD-Net optical server, each PC workstation had to load CD-Net memory resident drivers, which occupied about 30KB of conventional memory before loading a database application. This method of operation is commonly referred to as the "redirector," which is akin to the NetWare method of loading IPX and NETX shells to redirect all DOS requests through the network. Unlike NetWare shells, the CD-Net memory drivers were automatically unloaded to free needed memory. The second problem was that, at times, a mounted CD-ROM in the optical server could not be read. To correct the problem immediately, a librarian would manually reload the disc or reboot the optical server, which affected other users attached to the server.

With no technical support staff of its own, the research library requested help from LANL's Desktop Computing group to provide the needed expertise in PC/LAN management. After months of reloading disabled discs and reoptimizing workstation memory through QEMM optimizer programs, the decision was finally made to redesign.

CD-ROM NETWORK REDUX

By 1993, the redesigned CD-ROM network was beginning to take shape, but to keep costs down, a significant degree of ingenuity was required. The existing hardware and software components were carefully examined to determine what pieces could be reused for an upgraded CD-ROM network system. The objective was to create an open system, built with existing and off-the-shelf hardware and software that was upgradable, expandable, and scalable. With that objective in mind, the only

hardware used were the two tower cases and 28 CD-ROM drives from the CD-Net optical servers, as well as other spare parts.

Additional hardware and software products were purchased to manage the CD-ROM drives. They included Micro Design International's (MDI, Winter Park, FL) SCSI Express NLM-based software, selected for its ability to mount CD-ROMs as if it were a Novell mounted hard drive; a 486/66MHz EISA motherboard with 64MB of RAM to replace the 486/33MHz motherboard with 8MB of RAM in the NetWare file server; a 500MB IDE hard drive and controller, four 32-bit EISA SCSI host adapters (Buslogic 747) for controlling all 28 CD-ROM drives; a Standard Microsystems Corp. (SMC, Hauppauge, NY) 32-bit dual-channel network card; four external SCSI cables, and two internal SCSI ribbon cables.

Before the reconfiguration, the file server and two optical servers were separate units; now they are unified under the control of MDI's SCSI Express NLM. The CD-Net towers were also relabeled as SCSI Express towers. All CD-ROMs were easily mounted; each appeared with its own volume label and was mapped using the NetWare map command.

Those steps in creating a robust and expandable system helped eliminate the need for loading CD-Net's redirector device drivers at each PC station. It freed 30KB of memory per workstation, it also eliminated the need to reload the disabled CD-ROM drives under CD-Net. With the new system in place, problem reports on unreadable CDs dropped to zero, meaning reference librarians could now spend more time training and answering questions instead of fixing CD-ROM network problems.

Immediate benefits were realized under the SCSI Express NLM solution. Networking of ISI's Science Citation Index on CD-ROM behaved better in the new environment. Under the former CD-Net redirector solution, all CD-ROMs were split (in chronological order) into four menu options because of memory limitations—each mounted CD-ROM required 8KB of RAM of workstation memory. However, under the new SCSI Express method, all 14 CD-ROMs were simply mapped from drive `D:` through drive `Q:`, which required no additional memory. Science Citation Index was now accessible through one menu option, which pleased both patrons and library staff. Response time was perceived better under SCSI Express due to its use of NetWare's dynamic cache, as well. Each mounted CD-ROM under SCSI Express requires very little memory when used with the 64KB block of memory allowed under NetWare. For example, a CD-ROM with 633MB of data would only require 233KB of mount memory. If multiplied 28 times (the number of available CD-ROM drives), the file server would only need about 7MB of RAM.

COPY CAT

With the NetWare expertise now in-house, the next assignment was to build a RAID system onto which CD-ROM data could be copied, freeing up needed CD-ROM drives. The original intent was to copy all 14 Science Citation Index data discs onto magnetic media—after a hard drive copy license agreement was obtained. The custom-built RAID system would be an open, low-cost, modularized system built with off-the-shelf hardware and software components, as well as spare parts. The technical staff decided on a level 5 RAID system, which provides complete fault toler-

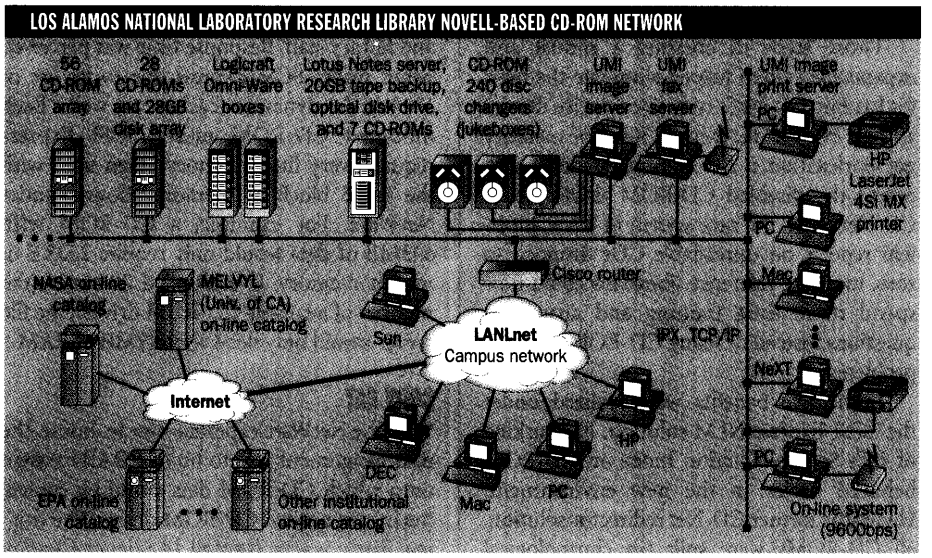


Figure 1. *The LANL research library's NetWare-based CD-ROM network provides access to journals in science, engineering, physics, chemistry, business, mathematics, geology, and medicine, and on-line databases.*

ance and the maximum amount of usable disk space within the array—between 85 percent and 93 percent of usable space compared with other RAID levels. A level 5 system provides data redundancy through distributed parity on a striped disk array, which simply means that if one drive fails, data can be recovered automatically by using the parity information on the other drives.

The RAID technology acquired for the project was the NLM-based Chantal (San Diego) Paragon Disk Array Software 4.5. High-end hardware components were purchased for building a robust custom-made RAID system which included a 486/66MHz EISA motherboard with 128MB of RAM; two 32-bit EISA SCSI host adapters (Buslogic 747); an SMC 32-bit dual-channel network card; five 2.6GB, half-height, Seagate Barracuda hard drives and five 2.4GB, full-height Micropolis (Chatsworth, CA) hard drives; two disk array cabinets (with internal SCSI ribbon cable) for housing the SCSI hard drives; one LAN server cabinet for the CPU motherboard, a 40MB IDE hard drive (for booting DOS) and VLbus IDE controller; and two external SCSI cables.

In December of 1993, the second NetWare 3.11, RAID 5 file server was born, with 16GB of on-line storage in which to copy and read CD-ROM data. The copying of Science Citation Index CD-ROMs began with the network copy command utility NCOPY, along with the /P parameter for preserving file attributes—hidden, system, and so on. The DOS XCOPY was used as well. However, some hidden and systems files could not be copied with those system utilities on a consistent basis. An alternative utility was required and the LAN administrator searched the Internet for the right tool. After downloading and testing several copy utilities, he found one that worked consistently: MCP.EXE, a freeware copy utility from Monk Software (Santa Clara, CA). Each copied Science Citation Index CD was assigned a subdirectory on the disk array; it was then root mapped with a network drive in order for the Science Citation Index search engine to see it as a CD-ROM drive. After copying all 14 discs onto the disk array, the data occupied the storage equivalent of only eight CDs (assuming a 660MB capacity for each).

As network services grew, more network hardware equipment was introduced, and by mid-1994, the floor space in the computer room became a critical issue. The crowded conditions were a potential hazard to non-staff maintenance workers, and to the equipment itself, which occasionally was found to be dusted with ceiling debris. To address the problem and secure the NetWare equipment, an industry-standard 19-inch, expandable rack-mount cabinet was purchased for the existing file server components, which included two CPUs, 10 hard drives, and 28 CD-ROM drives. Rack-mountable hardware enclosures were purchased for those components as was a CRT/keyboard AB-switch to toggle between the CPUs.

Benefits of the consolidated configuration included centralization, heat control, security, accessibility, mobility, and above all, a smaller footprint. After the server components were rack-mounted, the storage capacity of the RAID system increased from 16GB to 21GB. Later, another rack cabinet with 56 CD-ROM Toshiba XM-3401B drives was purchased for CD-ROM titles. NetWare 3.11 and SCSI Express NLM software were also installed on that system, which consisted of a 486/50MHz machine with 64MB of RAM, four Adaptec (Milpitas, CA) 2742 dual SCSI channels for controlling all 56 CD-ROM drives, and a pull-out tray with an LCD monitor and keyboard.

That completed, the library staff then turned its attention to the CD-ROM collection itself. Up until that time, all CD-ROM titles acquired had been science-oriented. It was decided that a business-oriented database was needed as well, to help researchers identify markets in supercomputing, lasers, superconductivity, materials research, environmental controls, and biomedical research. This type of resource would advance the laboratory's mission to seek U.S. industrial partners.

A network turnkey system from University Microfilms, Inc. (UMI, Ann Arbor, MI) was purchased for retrieval of business periodicals. Its flagship product, ProQuest PowerPages, is an image-based delivery system that provides digitally scanned images of articles exactly as they appeared in the original publications; a digital image can be printed or faxed. The system consists of three major components: an image server, an image printer server, and a fax server. The image server has three attached CD-ROM jukeboxes, each holding up to 240 CD-ROMs. When a user performs a search, an abstract of the article is presented; if an image is available, a copy of the image can be sent to either the image printer server or the fax server for output, but not to the screen.

DATABASE BENEFITS

The major benefit of CD-ROMs over on-line connection databases is that the laboratory's staff has access to the CD-ROM databases 24-hours a day, and seven days a week. Instructions on the use of each database are placed near each public workstation. On-line connection access is also available, but assistance is required due to the high cost of connect time. Reference librarians can help users combine the correct retrieval parameters.

On-line search requests and new CD-ROM titles are assigned to a research librarian according to his or her level of expertise in science, medicine, or business. After studying the assignment, the librarian prepares a training schedule for other laboratory staff members. These training sessions are conducted in a special training room equipped with six 486

PCs. Reference librarians must constantly learn new search engines as additional CD-ROM titles are acquired because there is no industry-standard search engine. Two library-standard search engines in use at the research library are SilverPlatter (Norwood, MA) SPIRS and Dialog (Rockville, MD) OnDisc, which have well-known interfaces in the academic marketplace.

The next objective was to create a universal PC workstation that would combine the functions of the public, training, and staff workstations. The universal workstation would have a Windows interface and allow access to Windows and DOS-based CD-ROM titles, e-mail (Eudora and Lotus Notes), and Internet access to other information resources through Netscape's World Wide Web interface. More importantly, it would facilitate cutting and pasting of information for staff research.

The Library Automation Team (LAT), comprised of six highly qualified technical personnel with expertise in PC, LAN, and Unix technologies, was assembled to support the growing technology environment. The LAT group gradually replaced the staff's PC workstations with Dell Pentium machines for processing document images and graphics on the Internet. The Saber LAN Workstation software for Windows (Sabre Software, Dallas) was chosen to provide ease-of-use, as well as to provide better metering and software usage statistics to determine CD-ROM title usage. Statistical data would allow librarians to determine if a CD-ROM title subscription should be renewed or replaced.

DIGITAL LIBRARY

The culmination of its efforts has resulted in a project the LANL refers to as the "Library Without Walls." Research library director, Rick Luce, says, "The conceptual paradigm changed from a focus on buildings, which house physical collections, to information services, which are bound neither by walls, nor traditional book and journal collections. In this new paradigm, library users connect remotely and use computer technology to access local and worldwide information providers—a concept known as the digital library."

The project's goal is to make digital collections available at the desktop. One challenging project looming around the staff is making CD-ROM titles available over the campus network, called LANLnet (see Figure 1), which means incorporating interoperability for multiple platforms and a mix of protocols to run DOS-based CD-ROM applications under those platforms. X Window has been chosen to remedy the problem and some hardware and software from Logicraft (Nashua, NH) has already been selected. Release of this service to the laboratory's patrons is imminent once all the details are worked out: site license restrictions, support issues, and standardization of keyboard and X-terminal server software.

CD-ROM technology is an important piece in a complex mosaic required to create the ideal virtual library. Other parallel projects will join in to create the total digital library objective. Explains Luce, "Dissemination of scholarly research in a timely manner to the researchers' desktops—whenever and wherever they need it—is the goal of the Library Without Walls project." ~

Lisandro R. Ramos is a technical staff member and an ECNE at Los Alamos National Laboratory. He can be reached via the Internet at lramos(?lanl.gov.