

# The Library That Isn't There

## Digital libraries transform books, photos, and videos into bits and bytes

By RICHARD LIPKIN

**T**he vision is grand: a world of literature, history, photographs, movies, and maps open, free of charge, to any curious mind that wants to meander through the electronic equivalent of library stacks.

Plug in a computer. Dial up a database. Enter a digital library. What might one find? A Sanskrit poem, a map of ancient China, and the U.S. Constitution, to be sure. Perhaps even plans for a new school, color pictures of local flowers, or clips from Fred Astaire films. One might even stumble onto segments of a documentary about the John F. Kennedy assassination or musical tracks from a recent jazz festival.

But the question that keeps computer scientists hopping is how to design such a system. What is the best way to organize, store, and link data in a worldwide "library" so that everyone from academic researchers to casual browsers can use it easily?

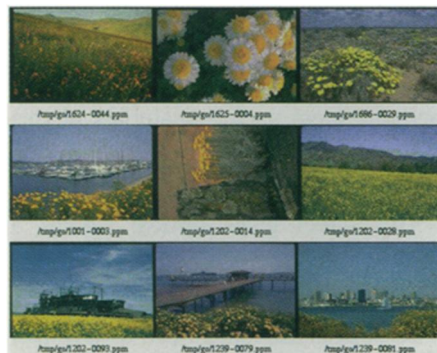
In a joint initiative to address this question, the National Science Foundation, the Defense Department's Advanced Research Projects Agency, and NASA have invested \$24.4 million in six projects. The goal of these projects, all centered at U.S. universities, is to find better ways of managing vast amounts of data in digital form. That information includes not only texts, but maps and charts, tables, sounds, spoken words, and video clips as well.

Once information has been transferred from paper or film to silicon storage, researchers must find easy ways to search through billions of bytes of material to retrieve desired items. The system must be simple enough for anyone to use.

These projects present more than just a technical challenge. For a digital library to be useful, its builder must take into account human behavior: why people go to libraries, what they look for, and how they select information. The typical library patron has only a vague sense of what he or she wants at the outset of a search. Only by asking and answering questions do most people focus their inquiries.

The process of assisting people as they decide what they really want to know and then helping them to find it requires practice. It is a kind of expertise that comes

easily to skilled librarians, whose experience computer scientists want to integrate into a digital library. "The processes that people go through to get information are very familiar to librarians," says Miriam A. Drake, director of libraries for Georgia Institute of Technology in Atlanta. "But so much is based on intuition. It's not well documented. This experience could avoid lots of wasted effort in designing digital libraries."



G. Ogilvie/Univ. of Calif., Berkeley

*Berkeley's digital library focuses on the California environment.*

**E**ach of the six large library initiatives will make its own contribution to the overall goal.

- At the University of Illinois at Urbana-Champaign, a digital library for engineers will make available journals, magazines, and scientific literature. To help users navigate the maze of megabytes, system designers have enhanced the search software from Mosaic, a widely used program for browsing databases on the Internet, to search collections. The Illinois system will give more than 100,000 people at 10 large midwestern universities remote access to full texts and pictures of tens of thousands of documents, says Bruce R. Schatz, a computer scientist at Illinois.

- In a similar vein, the University of Michigan at Ann Arbor will present a vast collection of sights, sounds, and words related to Earth and space exploration. This project will test, among other things, three types of "software agents," or sophisticated computer programs, to help people find what they want, according to Daniel E. Atkins, a computer scientist at Michigan.

The first agent interviews the library

user to determine what he or she wants to know and how deeply the person wants to dig for information. A second, "mediating" agent then coordinates simultaneous searches for the desired materials in several databases. Finally, specialized "collection" agents actually fetch texts, photos, sound bites, and video images from particular archives.

Students in Ann Arbor high schools and Stuyvesant High School in New York City will be among the first users to test the Michigan digital library. Computer scientists will then use the students' feedback to hone the system.

- At Stanford University, a digital library will link very different kinds of information sources into one integrated system, according to Hector Garcia-Molina, a Stanford computer scientist. The vast network will include not only books and journals, but data from private collections and scientific studies. To smooth over otherwise bumpy roads in searches, the system employs software enabling people to browse through disparate sources at the same time.

- Through the Alexandria Project at the University of California, Santa Barbara, anyone who wants to peruse maps, photos, satellite images, or graphs of Los Angeles, Ventura, or Santa Barbara will be able to pull them up on computer screens. This initiative aims to find more efficient ways to store and search through large numbers of pictures.

- Researchers for the digital library project at the University of California, Berkeley, are assembling a smorgasbord of environmental minutiae about the state. On-line, one finds pictures of local birds and flowers, as well as resource reports, county plans, aerial photographs, maps, videos, and sundry trivia. The system's designers are outfitting it with natural language and image processors to interpret users' requests and find the best pictures for them.

- Perhaps the most innovative digital library will have its home in Pittsburgh. Carnegie Mellon University and public television station WQED have teamed up to produce what they call the Informedia digital video library. Initially, it will contain 1,000 hours of videotapes from WQED's archives, video courses produced by the British Broadcasting Corp., and electronic

field trips created for public schools in Fairfax County, Va. The project seeks to encourage students to learn math and science by way of archived video documentaries and educational programs.

"Students can select portions of videos that interest them and watch the segments at home on their own desktop computers," says Howard D. Wactlar, a computer scientist at Carnegie Mellon. Part of the project involves learning how people interact with video clips. "Is searching through videos different from searching through texts?" Wactlar wonders.

The only way to find out, he says, is to watch people actually do it.

Students will begin testing a prototype system with 100 hours of video segments in late 1995, Wactlar says. Students and faculty from Carnegie Mellon and the Winchester Thurston School in Pittsburgh will provide feedback. The scientists will then tailor the digital library to students' actual behaviors and patterns of use.

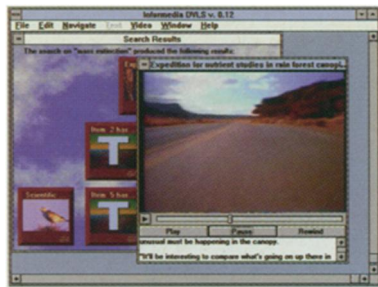
The Informedia system will employ a speech recognition program that automatically transcribes and stores video soundtracks as text. People can then search for and retrieve video segments on specific topics by choosing key words or by asking a question whose answer occurs in a segment. The system also apportions videos into distinct topical units, or video paragraphs, which can be viewed on demand.

Users of Carnegie Mellon's system can cruise through databases with voice commands. The system's builders are designing it to interpret language at a semantic level beyond the mere occurrence and recognition of key words. It will scan transcripts for topics that are implied, but not directly discussed, and analyzes video images to find scenic compositions, such as outer space or the deep sea or particular individuals who are speaking.

For instance, if someone asks the computer, "What problems are there with travel to Mars?" the system might call up a video interview with science correspondent John Noble Wilford of the New York Times. "Ultimately, the venture is probably going to be an international one, because of the cost," he says. "We're talking about hundreds of billions of dollars over many years."

When asked, "What does Arthur C. Clarke think about a Mars mission?" the computer would fish out of its archives an interview with the science fiction author. "Any Mars exploration will be international, if only because of the time frame involved and the money," Clarke tells his audience. "We're not going to have an expedition to Mars until the next century. By that time, there may not be any nations as we know them now. There may be hundreds of tribes."

Of course, since there's no such thing as a free lunch, especially in the computer world, scientists will test pricing and billing techniques for use of library



Carnegie Mellon's Informedia project sports digitized video clips.

K. V. Pendyala/Carnegie Mellon Univ.

resources. One method, called NetBill, offers a way to charge for transmitting small quantities of information over a network. Privacy and security mechanisms to protect both users and archived materials will also undergo testing as part of the Informedia tryout.

**P**reservation of rare books and deteriorating manuscripts is another reason to put print into electronic form. Cornell University has put more than 2,000 19th- and early 20th-century books into electronic format. Each book, its acidic paper crumbling, is scanned page by page and stored on compact discs. The discs, each of which holds 20 books, are racked up for viewing in digital jukeboxes.

Readers of electronic materials cannot lose or damage them, says Anne R. Keney, a librarian at Cornell. Patrons of digital libraries, she adds, "will never face an overdue fine."

Preservation of crumbling books speaks loudly to the conservators of the Library of Congress, the British Library in London, and the Vatican Library—all of whom are digitizing their most treasured volumes. The Library of Congress' American Memory project now has more than 210,000 items in electronic form. Ready to go on-line are historical texts and the Leonard Bernstein archives. By the year 2000, the library expects to have digitized 5 million books and images.

The British Library, with more than 18 million volumes, has put its entire card catalog on line, going all the way back to "the beginning of printing," says librarian Jonathan Purday. An electronic Photo Viewing System already holds more than 10,000 historical images, ranging from Victorian spiritualist paintings of ghosts and seances to manuscript illustrations and baroque sheet music. More than 1 million international patents are stored on 16 literary jukeboxes, each with 100 discs. An electronic *Beowulf* project has preserved the original manuscript of the 11th century Anglo-Saxon epic.

In Rome, the Vatican does not want to risk losing historical manuscripts. An effort to put many of its treasured tomes into digital form will permit someone dialing in to peruse, for instance, the four oldest known manuscripts of Virgil's

poetry or a hand-illustrated version of *The Divine Comedy*.

**T**he process of designing and building digital libraries essentially from nothing has raised many unexpected questions for researchers. For example, what is a library? Is it fundamentally a building, a collection of books, or a way of organizing materials?

Moreover, what social functions do libraries serve? Should libraries aim chiefly to disseminate information as widely as possible or to protect, preserve, and conserve it?

Traditionally, Western culture has accorded the mighty book storehouses a certain status, style, and function. But new technologies have called the classical view into question. For instance, a national system of digital libraries may not fill even one grand marble edifice. In an electronically oriented society, library patrons may focus on access to digitized pictures and texts rather than places to go to read or borrow books.

"One way to look at libraries is to see them as a knowledge base for society," says Terence R. Smith, a computer scientist at Santa Barbara. "Designing digital libraries gives us an opportunity to redefine what a library is without losing the great resources that libraries already are."

Smith contends that taking a fresh look at what a library could be may prove instructive. For instance, he says, a digital library might one day be a place where a geologist could obtain and process data about a region to help predict the likelihood of an earthquake. It might be somewhere a businessman can go to find the income distribution, office options, and parking availability of a potential development site. It might also be a way for a third grader, at home with a computer, to watch a documentary about the life and times of Amelia Earhart and print out a chart of Earhart's final flight.

"In a traditional library, these types of information are scattered around," says Smith. "The question now is how can we bring them all together in an easily accessible system?"

For starters, says Robert Wilensky, a computer scientist at Berkeley, digital library designers should think less narrowly about the kinds of services libraries can perform. For digital libraries to succeed, their designers ought to abandon traditional notions of libraries altogether, says Wilensky.

"Maybe in some deeper sense, libraries as buildings have functioned as intermediaries," Wilensky adds. "But now the information that used to be bundled up in a building will be distributed all over the place. So libraries as physical places filled with objects may become obsolete."

"A digital library will be just like any traditional library," Wilensky quips. "Except that there's no library there." □