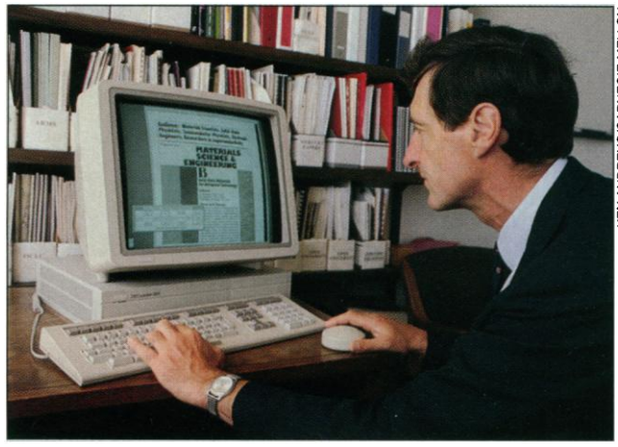
Virtual Libraries, Complete With Journals, Get Real

Libraries are not favored haunts of chemists, or many other research scientists. Rather than spending hours in vain pursuit of elusive citations, they personally subscribe to—and clutter their offices with—about half a dozen core journals. They want first crack at the hottest papers before the competition gets to them, and while libraries subscribe to the same journals the copies could be missing, loaned out, or damaged. But what if those journals were available through the researcher's desktop computer, which could scan not just six but hundreds of journals and automatically retrieve the relevant articles? And what if those articles were more than just text on a computer screen, but complete copies with detailed graphs and pictures?

That type of library might be haunted more often. And it's virtually here. On 21 April, the Chemistry Online Retrieval Experiment, or CORE, a "virtual" library, premiered at Cornell University. It offers Cornell chemists 20 American Chemical Society (ACS) journals—not abstracts, not text, but entire articles with pictures, tables, graphs, and captions conveniently displayed on office computers. A chemist can type in, say, "temperature, reagent, concentration," and a list of articles appears. A quick click with a mouse on one item and a new window opens with just the graph he or she needs from *Analytical Chemistry*.

The CORE project is only one of several new virtual library programs. Others are under way at Carnegie-Mellon University (CMU), the Naval Research Laboratory, and Columbia University. To make the virtual into reality, engineers and programmers have surmounted one major obstacle: designing systems capable of handling the huge volume of graphic material—charts, tables, pictures of gels—that is an essential part of an average scientific journal. But whether those systems are fast enough to be easy to use, whether directed electronic searches will rob science of spontaneous discoveries gleaned from literature browsing, and whether journal publishers will adapt to systems that could eat into some profits, remains to be seen.

CORE, the most advanced of the current wave of virtual library projects, is a joint venture involving Cornell, ACS, Bellcore, and a database searching service called the Online Computer Library Center (OCLC). CORE currently contains ACS journals from January 1991 through September 1992; that's approximately 20,000 articles, or 142,000 pages. The collection eventually will go back to 1980 for most journals, and would otherwise take up more than 600 feet of shelf space. Getting that mass of data online in a form that's easy to use means some extreme complexity behind the screens. The journals are produced by the publisher with software that sets the appearance of the printed page. That raw data—formatting codes and text and tables and small graphics—is recoded so it can be displayed on a desktop monitor and then sent to the library. There, a user works with a graphical, Windows-like interface developed by Bellcore to activate OCLC's search program, which scans the data for key words a user types in using familiar "and-or" search strings. A list of matching articles pops up in a window, the user chooses a title, and images of the article appear on the screen.



KEN ANDREYO/CARNEGIE-MELLON

Picture this. A researcher peruses a journal formed electronically in Carnegie-Mellon University's virtual library project.

Since detailed graphics take up huge amounts of memory, at CORE they are scanned in separately and kept as compressed data files on optical storage disks in a mechanical "jukebox." The disks can be flipped through and their images called to the screen at the click of a screen button. The Cornell chemists' window into this virtual world is a Unix workstation or a Mac IIci with MacX software, Ethernet cards, 5-6 Megabytes RAM for System 7, and at least a 68030 processor.

What excites the average user, however, is not the gadgetry but the newfound ability to get the big picture. "It's like having *Chemical Abstracts*, except you get the whole paper," says Cornell theoretical chemist William Shirley. And being able to get the whole paper online could revolutionize the way scientists keep up with the literature. Keeping up is, after all, the reason "most faculty subscribe to five or six core journals personally. That way they know that nothing in those will slip by them," says Jan Olsen, the Cornell library director. "But CORE will bring a journal to them when they want it, where they want it, and as often as they want it, whether or not someone else is using it. They will tell the computer to 'search this' and the computer does it very rapidly. It's a better job of keeping up than they can do if they have to physically eyeball all the literature every time they want something."

One Cornell graduate student grateful for less wear and tear on her eyes is Kimberly Lawler, who was thinking about doing a study on carbonyl coupling. Not knowing precisely where to look, Lawler used her CORE training session to search for "McMurry reaction," "carbonyl," and other terms and came up with a full screen of references. "You can do it in your office and see if it's right for you," she says.

Lawler's endorsement is not wholehearted, however. "Big articles online are a pain because it's harder to flip through the pages," she says. It's true that new pages in an article don't spring to the screen instantaneously, like turning the pages of a book. The mechanical jukebox that holds the complete images slows things down while it flips to the proper disk, and optical drives are inherently slower than magnetic drives. It also takes time for stored images to be decompressed and transmitted across the network. All in all, a new page can take several seconds to appear while the researcher fidgets in his or her seat.

Speed will doubtless improve with better hardware, as it has throughout the history of computers. Other aspects of CORE might prove more troublesome. The efficient search routines, for instance, might cut down on scientific serendipity. "[Scientists] often approach the literature with no specific idea in mind," Olsen says. "They are not looking for a particular author or title, or compound or concept; they are simply looking for an intellectual adventure. It stimulates ideas, research paths, that are completely unpredictable to them. You can't do that as well with the electronic text. I think this will get them frustrated."

Not everyone agrees with this, and some scientists discount it entirely. It is possible, after all, to browse whole issues electronically by assembling a Table of Contents from an issue date. A mouse click on a line will bring up the article. "You can browse this system just like the printed page," says Rich-

ard Entlich, senior systems analyst. "Searching is actually very visual."

You can also print out what you've found. That may be nice for the scientist but it is one aspect of virtual libraries that makes publishers quite nervous. They worry about bootleg journal copies cutting into their circulation rates, and they're not sure how libraries will keep these virtual journals from becoming part of a pirate publishing industry.

Those issues are of particular concern at the Cornell project, which is being carried out in concert with ACS, the journals' publisher. The same situation applies to another online library project called TULIP, or The University Licensing Program, which is taking shape with the help of the journal publisher Elsevier. TULIP is furthest along at CMU, where researchers have access to seven Elsevier artificial intelligence journals plus the *IEEE Computer Journal* and CMU computer science monographs.

Cornell, CMU, and other institutions are considering several different ways to protect the rights of the journal publishers. Restricting laser printing of journals within the institution by proper authorization and validation (computer IDs or library card) may be one way to solve the problem. And some pay-for-use systems will probably emerge, especially if the library offers access to journals to which it doesn't already have a subscription.

Printing out articles means one other thing: Virtual libraries won't, alas, eliminate the cluttered office. "We do not have strong evidence that people want to leave their hard copies," says CMU's head librarian Charles Lowry. "You put up a Novell token ring and what do people do? They print out their e-mail!"

Nobody really knows for sure how people are going to respond to online reading until the systems are up and running for some time. Users may go to reading off the screen, or they may print and then read. It's hard to resist having something to hold in your hands—just ask Cornell's Entlich. Searching his desk for handouts, he throws up his hands and exclaims, "In the midst of an electronic library project, I'm buried in paper!"

–Larry Krumenaker

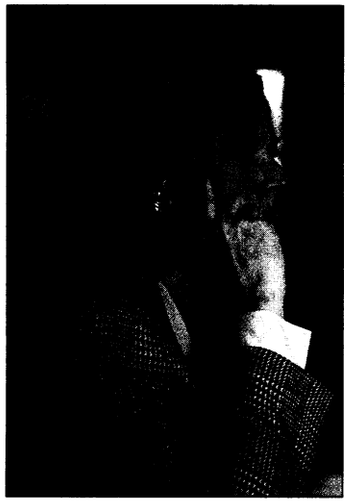
Larry Krumenaker is a science writer based in New Jersey.

BIOMEDICINE

Healy Slams Clinton's NIH Budget

Bernadine Healy may be on her way out as director of the National Institutes of Health (NIH), but she is not leaving quietly. Last week, with just over a month left in her tenure, Healy lashed out at the Clinton Administration for effectively cutting the NIH budget.

The venue was the House of Representatives Appropriations Committee, which, along with its Senate counterpart, will ultimately set the agency's 1994 budget. In her final appearance before the committee, Healy quickly departed from her prepared remarks, which had been approved by the Administration, to blast Clinton's 1994 request for NIH. "I am deeply troubled that in 1994, NIH faces a budget that is a contraction in virtually every category," she said. If Congress cannot substantially increase the Administration's $10.67 billion request for NIH, the word "devastating," she said, "would be mild."



RICK KOZAK

Not fading away. Outgoing NIH Director Bernadine Healy.

The problem, Healy said, is that the budget is considerably worse than it looks. Taken at face value, it gives NIH a 3.2% increase over last year. But most of the new money, Healy argued, is actually accounting tricks and funding for a few favored programs. As an example of both, she cited a special $216 million breast cancer research request. Designed to continue a program that Congress had for political reasons placed in the Army's accounts last year (although the 1994 request is for new money, and does not affect the Army program), it bears the traces of its military roots: Like many Defense Department programs, it is "forward funded," meaning that NIH must distribute the money over the average 4-year life of the grants it will fund. NIH can spend only one-quarter of the money in 1994.

Breast cancer isn't the only research area that makes the budget picture seem brighter than it is, said Healy. The Administration has also requested $214 million in research on such topics as high-performance computing, advanced materials, and health-care reform, none of which has been an NIH focus before (*Science*, 23 April, p. 483). After you have accounted for all these favored programs, the core basic research programs at NIH are actually due for a 1% cut, Healy said. Factor in the biomedical inflation rate, and the cut grows to slightly more than 4%. As a result, Healy said, NIH has decided to eliminate the usual cost-of-living increases for its grants next year (which left her "very distressed") so that the agency can keep the number of new and competing grants at about this year's level of 5600.

Congress, which has traditionally rescued NIH from budget predicaments, may not be able to offer much help this year. In fact, it may make things worse. "The president's budget could be the high-water mark," says David Moore, assistant vice president for government relations at the Association of American Medical Colleges. The reason: With Clinton's budget still some $8 billion over the deficit-reduction targets set by Congress, something will have to give when Congress splits up the budget pie among the various appropriations committees next week. The most likely casualty, according to an appropriations committee staffer, is the allocation for the Labor, Health and Human Services, and Education committee, which, besides including the NIH spending, has the largest single share of the federal budget.

Healy had several explanations for NIH's troubles. Early in her testimony, she pointed the finger at the Congressional Budget Office, which in its February report listing options for deficit reduction suggested that a cut in NIH's budget could "be justified by its rapid growth in recent years." And Healy complained that NIH "doesn't have a seat at the table" to head off such threats.

For example, Healy said that neither she, nor any other senior NIH staff, has been asked to participate in the topic of the month: health-care reform, which may dictate budgets for years to come. Nor has there been a champion for biomedical research in NIH's parent agency, the Department of Health and Human Services (HHS), or at the White House, according to HHS officials. Philip Lee, the incoming HHS assistant secretary for health, could be such a champion, but he has not yet been confirmed by the Senate.

All that may simply be the price of transition. Indeed, HHS Secretary Donna Shalala, testifying at an earlier appropriations hearing last week, hinted that NIH may fare better once a new director is in place. She said she was "reviewing recommendations" for NIH which may, HHS sources say, include more power for the NIH director and a greater role in administration priority-setting. This may bode well for 1995, but neither Shalala, nor anyone else for that matter, was able to offer anything to temper what looks to be a dismal 1994 for NIH.

–Christopher Anderson