Beyond Databases and E-Mail

In the coming decade, computer networks such as the Internet will link researchers into "electronic communities" that will create new ways of collaborating and sharing information

News Reports: Computing the Future

No longer specialized number-crunchers, computers have changed the lives of scientists from every discipline in ways too many to count. Electronic mail has transformed scientific communication, electronic databases have expanded access to information, and word processing programs have eased the writing of manuscripts. But by all accounts the computer transmutation of science has only just begun. In the news component of this special section, *Science* takes a look at some of the coming hot spots —networked research communities that erase geography, schemes for handling massive amounts of data from Earth-observing satellites, imaginative new uses of artificial intelligence, and a software approach that puts powerful programming tools in the hands of nonexperts.

Editorial

Perspectives

Articles

Editors: Robert Pool, News; David Voss, Perspectives and Articles

Robert Weller has a dream. The Woods Hole Oceanographic Institution oceanographer dreams that one day he will be able to turn on his computer in the morning and transform his office into a global oceanography institute. In his dream, he can bring up data on water temperature, air temperature, wind speeds, currents, air pressure, humidity, salinity, and solar radiation from ocean monitoring devices all across the world —all there at his fingertips and all as fresh as today's catch. Then, without leaving the keyboard, he can analyze this data, borrowing time on a supercomputer that may be halfway across the country. If he wishes, he can link up with colleagues from other institutions to look for patterns in the data and compare the details with predictions from ocean models. Results in hand, he can then instruct the monitoring devices to modify their observations to answer new questions or improve later data.

A dream, yes. Today, scientists like Weller can use computer networks to pass E-mail or retrieve archived data from databases, but the sort of instantaneous give-and-take he envisions is just not possible. That's changing, however. The tools that Weller dreams of—and many more—are on the horizon for scientists in all fields, thanks to the ongoing explosion in computing power, ever-growing communications capabilities, and entirely new ways of thinking about what computers can do. It's impossible to predict exactly what shape the future will take, but if the computer gurus are correct, at least one forecast can be made with confidence: Over the next decade or two, "electronic communities" like the one in Weller's dream will spring up, linking scientists much more closely than ever before with information, instruments, and far-flung colleagues. "There is genuinely a revolution here," says Bruce Schatz, an information scientist at the University of Illinois. "Life is going to be completely different."

The foundation for that transformation has already been put in place. The Internet, the worldwide system of interconnected computer networks, has created an environment ripe for revolution. There are now 1.7 million host computers hooked up to Internet worldwide (a million in the United States alone), with somewhere between 5 and 15 million individual users, says Vinton Cerf, vice president of the Corporation for National Research Initiatives in Reston, Virginia, and president of the Internet Society. And the numbers are doubling each year, he says.

Although there are no statistics on what percentage of scientists are Internet users, anecdotal evidence suggests that some communities, such as computer science and electrical engineering, have almost 100% participation. Even the humanities, traditionally leery of high-tech gadgetry, are joining in. "There's a sort of virtual university being created [with electronic mail]," says James O'Donnell, professor of classical studies and coordinator of the Center for Computer Analysis of Classical Texts at the University of Pennsylvania. "There's one colleague [at another university] whom I used to speak with maybe a couple of times a year. Now we're in e-mail contact twice a day."

Indeed, electronic mail is often the major reason a researcher signs onto Internet, Cerf says. But once there, he or she finds plenty of other uses. In physics, for example, Paul Ginsparg at Los Alamos National Laboratory has set up several bulletin boards that accept preprints and send out the abstracts to thousands of subscribers, who can download full texts of papers that seem interesting (*Science*, 26 February, p. 1246). In biology, researchers routinely dial into hundreds of databases containing genetic maps, protein structures, and so on (*Science*, 11 October 1991, p. 201). All told, there are now approximately 50,000 databases available over Internet, Cerf says, with many more coming on line each month.

But this embarrassment of riches has created a problem: How can you find anything in that mass of data? The database file names are usually nondescriptive, so unless you know what you're looking for, you probably won't find it. Even worse is the difficulty of learning which databases are out there and what information they contain. It's as if you have wandered into a library full of books, but the books are identified only by number, and when you open one to its table of contents, the chapter headings are equally unhelpful.

What's needed is better data about the databases, and that's where the first stage of the revolution is occurring. Over the past 18 months, the Internet library has begun to supply its books with "labels" and "tables of contents" and even to offer indexes of its

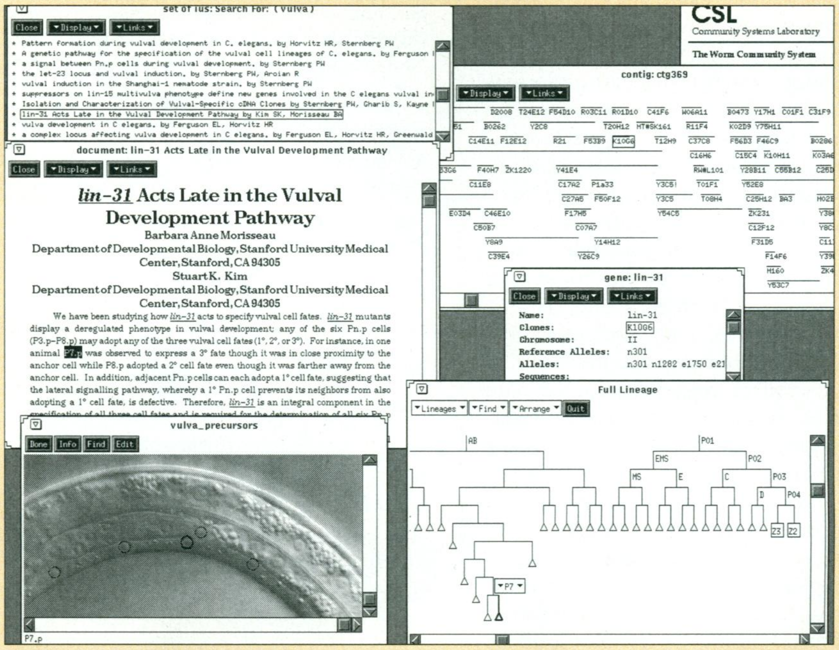
Networking the Worm

Caenorhabditis elegans is an unprepossessing critter, a millimeter-long soil-dwelling worm that eats bacteria. Until recently, its main claim to fame has been that developmental biologists, who have adopted *C. elegans* as a model organism, understand its development more completely than that of any other multicelled creature. They have mapped out exactly where each of the worm's 959 cells lies and when each cell appears during the 3-day passage from fertilized ovum to fully functioning adult. Now, however, the tiny, transparent roundworm can add a second distinction to its c.v.: It is the subject of the most sophisticated and ambitious computer information network yet created, a network that offers a preview of what scientists in other fields can expect to encounter in coming years (see main text).

The Worm Community System (WCS) can be thought of as a "hyperlibrary," says Bruce Schatz, the University of Illinois scientist who developed it. At its heart is a computer network linking many *C. elegans* databases, formal and informal, so that a researcher can retrieve related information from many databases at once. The formal databases include gene descriptions, maps of the worm genome, DNA sequences, and journal articles—in effect, just about all "official" scientific knowledge on the worm. The informal databases fill in the gaps with such things as notes on experimental methods, lists of researchers, and the text of the *Worm Breeders Gazette*, the community's informal newsletter.

Much of this information was available through the Internet before Schatz appeared on the scene, but it was in isolated databases. A researcher would have to enter one database to retrieve a physical map, a second to pull out the DNA sequence of a particular gene, and a third to look for relevant literature on that gene—a clumsy, time-consuming procedure.

Schatz's scheme changes all that. Suppose you're interested in genes involved in the worm's sense of touch. You begin by entering "sensory," and the system finds every piece of literature that contains that word, displaying a one-line summary of each. Next, you perform what Schatz calls a "group follow" to get all the genes mentioned in that literature. Each gene—or any set of genes—can then serve as the starting point for a new search. A user might, for example, ask for a display of a genetic map indicating the locations of all these genes. Or the user can choose one gene and get its sequence, its location on a physical map, or a list of genes that have related functions.



set of ius: Search For: (vulva)

Close ▾Display▾ ▾Links▾

- Pattern formation during vulval development in C. elegans. by Horvitz HR, Sternberg PW
- A genetic pathway for the specification of the vulval cell lineages of C. elegans. by Ferguson
- a signal between Pn.p cells during vulval development. by Sternberg PW
- the let-23 locus and vulval induction. by Sternberg PW, Aroian R
- vulval induction in the Shanghai-1 nematode strain. by Sternberg PW
- suppressors on lin-15 multivulva phenotype define new genes involved in the C elegans vulval in
- Isolation and Characterization of Vulval-Specific cDNA Clones by Sternberg PW, Gharib S, Kayne
- lin-31 Acts Late in the Vulval Development Pathway by Kim SK, Morisseau BA
- vulva development in C elegans. by Ferguson EL, Horvitz HR
- a complex locus affecting vulva development in C elegans. by Ferguson EL, Horvitz HR, Greenwald

document: lin-31 Acts Late in the Vulval Development Pathway

Close ▾Display▾ ▾Links▾

lin-31 Acts Late in the Vulval Development Pathway

Barbara Anne Morisseau
Department of Developmental Biology, Stanford University Medical Center, Stanford, CA 94305
Stuart K. Kim
Department of Developmental Biology, Stanford University Medical Center, Stanford, CA 94305

We have been studying how *lin-31* acts to specify vulval cell fates. *lin-31* mutants display a deregulated phenotype in vulval development; any of the six Pn.p cells (P3.p–P8.p) may adopt any of the three vulval cell fates (1°, 2°, or 3°). For instance, in one animal P7.p was observed to express a 3° fate though it was in close proximity to the anchor cell while P8.p adopted a 2° cell fate even though it was farther away from the anchor cell. In addition, adjacent Pn.p cells can each adopt a 1° cell fate, suggesting that the lateral signalling pathway, whereby a 1° Pn.p cell prevents its neighbors from also adopting a 1° cell fate, is defective. Therefore, *lin-31* is an integral component in the specification of all three cell fates and is required for the determination of all six Pn.p

vulva_precursors

Done Info Find Edit

P7.p

CSL
Community Systems Laboratory
The Worm Community System

contig: ctg369

▾Display▾ ▾Links▾

gene: lin-31

Close ▾Display▾ ▾Links▾

Name: lin-31
Clones: K10G6
Chromosome: II
Reference Alleles: n301
Alleles: n301 n1282 e1750 e21
Sequences:

Full Lineage

▾Lineages▾ ▾Find▾ ▾Arrange▾ Quit

D2008 T24E12 F54D10 R03C11 R01D10 C41F6 W06A11 B0473 Y17H1 C01F1 C31F9
B0262 Y2C8 T20H12 MT#SK161 R11F4 K02D9 Y75H11
C14E11 F12E12 R21 F53B9 K10G6 T12H9 C37C8 F56D3 F46C9 B0286
C16H6 C15C4 K10H11 K03A6
F40H7 ZK1220 Y41E4 RW#L101 Y28B11 C55B12
C11E8 C17A2 P1a33 Y3C5! T01F1 Y52E8
C27A5 F50F12 Y3C5 T08H4 C25H12 BA3
E03D4 C46E10 F17H5 Y54C5 ZK231
C50B7 C07A7 C12F12
Y8A9 Y14H12 F31D5
C39E4 Y26C9 F14F6
H160
Y53C7
AB
P01
EMS
P02
MS
E
C
P03
D
P04
Z3
Z2
▾P7▾

Windows on the worm. A search through the Worm Community System pulls up various kinds of information from databases scattered around the country.

BRUCE SCHATZ

This ability to jump from one database to another depends on "links"—software connections between different pieces of information. Gene descriptions in one database, for instance, are linked to the locations of those genes on a physical map in another database. Schatz had to create the original links himself to get the WCS started, but ultimately it will be up to researchers using the system to add new links.

Users would create these links, as Schatz pictures it, each time they add their own information to the system. Suppose a researcher enters some data on a new gene, and he has noticed a functional similarity between his new find and a known worm gene. He can create a link between his gene and the older one and add a note explaining the similarity between them. The next time a scientist examines one of the genes and asks for related objects, the second gene will pop up.

"The way you get this giant interconnected space is to let the community do it," Schatz says, with everyone freely sharing information and ideas about how the information relates to what is already known. In this way far-flung researchers become part of a close-knit information community, building knowledge about a subject in much closer collaboration than was possible before.

But for that to happen, WCS has to catch on among worm biologists—and that's happening only slowly. Since it was set up 2 years ago, 25 of the 100 or so major worm labs around the world have signed on, and Schatz says another 25 have indicated that they plan to begin using it over the next year. At the Institute for Genomic Research in Gaithersburg, Maryland, Chris Fields says he finds Schatz's system particularly useful for browsing: "If I'm interested in this or that particular gene, then I look around." But like most other worm researchers, he says, he has used WCS only to get data out—not to put in new information to share with other scientists.

Fields says that's partly because creating new links was inconvenient in the first version of the software. A second version, released a few weeks ago, should lead to more use of this interactive function, says Schatz. But he thinks there's another reason for the slow start. The system offers "a genuine revolution in how people are going to interact with knowledge," and it's going to take a while for researchers to get used to the idea of being members of an "electronic community," freely sharing their knowledge with other scientists in an ever-growing hyperlibrary. Nonetheless, the WCS shows enough promise that the recent NRC report on "national collaboratories" pointed to it as a prototype for future electronic communities. Today *C. elegans*, tomorrow the world.

–R.P.

holdings. Driving this transformation are independent researchers who have developed software tools to help navigate the Internet and then distributed them for anyone to use, says George Brett, director of the Clearinghouse for Networked Information Discovery and Retrieval in North Carolina. Now Internet devotees say they don't know how they ever lived without such aids as Gophers, Wide-Area Information Servers, WorldWide Web, Archie, Veronica, and Jughead.

Archie, developed by graduate students at McGill University in Montreal, was one of the earliest of these tools. In essence, Brett explains, Archie is a guide to databases—a regularly updated index of the files available at various computers attached to the Internet that can be downloaded by any Internet user. "Ask Archie, 'Where is Kermit software?' and it will tell you all of the thousands of places to find it."

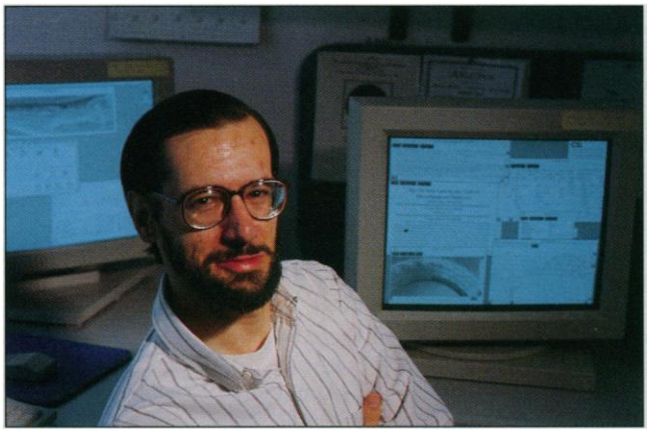
Archie is aimed at people who know exactly what they're looking for and simply need help tracking it down. But the most popular Internet aid, Gopher, is what Brett calls a "browsing tool." Gopher, created 2 years ago at the University of Minnesota, provides an easy-to-use gateway into "infobases"—generalized databases that can contain not only data but any other type of information, including text and multimedia displays. Installed on an individual infobase, the Gopher software supplies a nested table of contents, so that users can zero in on desired information by making a series of choices. Someone interested in New Zealand, for instance, can connect with a Gopher supplied by the Wellington, New Zealand, city council, pulling out descriptions of culture, geography, or whatever tickled his fancy. To date, some 1100 infobases inhabit "Gopherspace"—a virtual library, available publicly through the Internet, consisting of infobases that can be searched via a Gopher —and the number is growing rapidly.

The ranks of the Internet aids also include Jughead and Veronica. Jughead allows a user to search directly for files in an infobase that is indexed by a Gopher, while Veronica can look for files across all Gophers—it is an "Archie for Gopherspace," Brett says.

The most ambitious of the Internet tools to date is WorldWide Web, which Brett describes as a "hypermedia browser" for a collection of databases. Modeled after hypertext programs, it allows a mouse-equipped user looking through one file to move immediately to a related file in a completely different database by clicking on a highlighted word or phrase. In this way a user can trace a chain of ideas or information through a series of files that appear in computers around the world without worrying about where each piece is located. WorldWide Web has recently moved out of its demonstration stage, Brett says, and is being put to "real work" with some 64 databases included to date.

This trend toward making more and more information easier and easier to get will certainly continue, but a still bigger change is looming, the experts say. Increasingly, the Internet will be used to create "electronic communities"—collections of researchers in a single field who are linked electronically and who share information, instruments, software, and even computing capability.

Virtual laboratories. Some of the first such communities may be the "national collaboratories" touted in a recent report by the National Research Council (NRC).* A collaboratory would, as its name implies, integrate people and resources in such a way that a researcher in any location could hook into the system and do his work as if everything he needed—data, computing power, software, instruments, even other researchers—were right in the same building. "You have on your desk all the tools you need for a class of problems and they're all integrated together," explains William Wulf, professor of computer science at the University of Virginia and originator of the "collaboratory" term.


TOM IVES

Electronic visionary. Bruce Schatz thinks electronic links will trigger a revolution in "how people interact with information."

This sounds very much like the dream of Woods Hole's Weller, as well it should, since the NRC report pointed specifically to oceanography along with space physics and molecular biology as fields that could greatly benefit from collaboratories. The three diverse subjects share a feature that makes them good candidates for electronic collaboration: Each has its vital information spread out over many institutions.

Space physicists, for example, collect a wide variety of data in trying to understand how the sun's radiation interacts with the atmospheres of Earth and other planets. "Historically, these measurements have been sent to different institutions and then the people involved must come together and bring their data to compare," says Christopher Russell, a space physicist at the University of California, Los Angeles, who served on the NRC panel. A space physics collaboratory would give researchers access to all of this data electronically, along with the tools —computers and software—necessary to analyze it. It would also allow scientists at different institutions to hook up over the system in order to examine and manipulate the data together.

Eventually, predicts Schatz, another member of the NRC panel, such electronic communities will revolutionize "how people interact with information." To Schatz, the most important part of these communities will not be the fact that researchers can control instruments remotely or access data from dozens of sources at once, but rather that the communities will create a new way for scientists to record and share information and insights. In a prototype community Schatz developed for molecular biologists who study the nematode worm *C. elegans* (see box), scientists can not only add their own data to the system's databases but also create "links" between different pieces of information. Thus a researcher who notices, say, a similarity between two genes can leave a record of that observation; anyone who later brings up information on either gene will be informed of the similarity. This linkage ability, Schatz predicts, will turn computer networks into something much more than the high-powered library and communications systems they are today.

Since the software is developing so rapidly, the only factor limiting how quickly these electronic communities develop is likely to be scientists' willingness to take advantage of their features, Schatz says. When the telephone was invented, he says, "people wondered why they needed it when they already had the telegraph." With that kind of cultural inertia, electronic communities may not come into their own until people move beyond exchanging messages and scanning databases—things that could be done (albeit more slowly) by telephone, fax, and overnight mail—and start using the networks to collaborate in new ways. "The people who are carrying out this revolution," Schatz says, "are the graduate students," those budding researchers who have grown up with the personal computer and who aren't wedded to old-fashioned ways of doing science. They'll be the citizens of tomorrow's electronic communities, inhabiting a world that to Robert Weller seems much like a dream.

–Robert Pool

*"National Collaboratories: Applying Information Technology for Scientific Research." National Academy Press, 1993.