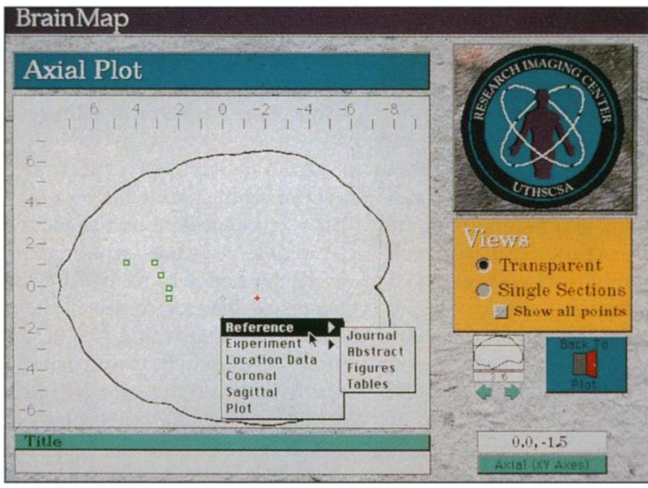
Databasing the Brain

Neuroscientists aim toward a confederation of databases that would let researchers wander through the brain from molecules all the way up to function

At the top of the brain is a bundle of fibers that poses a long-standing challenge to neuroscientists. The challenge is quite basic: Nobody knows what the front third of this bundle, the cingulate gyrus, does. So, when Washington University neuroscientist Steven Petersen attended a workshop* on databases recently, the cingulate gyrus struck him as just the thing for putting a prototype neuroscience database through its paces.

Petersen sat down at the keyboard, punched in the structure's name, and, in short order, the database (dubbed BrainMap) coughed up five studies in which human subjects performed complex motor tasks while the activity in their anterior cingulates was monitored by positron emission tomography (PET). A few minutes later, the screen showed an outline of a human brain with a graphic summary of how the parts of the cingulate seem to be involved in motor tasks. Petersen, for one, was impressed: "If we had all the data in one place like this, maybe we could figure out what this mysterious area is doing."



UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER

Data template. BrainMap's screen display.

That hope—that having all the data in one place will shake loose new insights about how the brain works—is driving a surge of interest in neuroscience databasing. Last year, the Institute of Medicine (IOM) enthusiastically endorsed what it called the "Human Brain Project," an ambitious, two-decade effort to develop a set of neuroscience databases (*Science*, 28 June 1991, p. 1794). And when the National Institute of Mental Health (NIMH) held an all-day workshop on the project at the Society for Neuroscience meeting in Anaheim in October, 400 neuroscientists crowded in, and more were turned away at the door. Indeed, the vision of a neuroscience community linked by databases is already becoming a reality: Several progenitor databases, including BrainMap, are under development and will compete early next year for millions of dollars in federal funding set aside for pilot studies for the Human Brain Project.

In one sense, neuroscientists are simply following the lead of the geneticists, protein modelers, and molecular biologists who have already gone on line to share a single, relatively standardized kind of data with their colleagues. But California Institute of Technology neuroscientist Jim Bower, who has developed Genesis, a three-dimensional simulator of brain circuits, argues that "what neuroscientists are trying to do is orders of magnitude more complex." He and other database visionaries see BrainMap and its like ultimately growing into a network spanning the entire range of subfields in neuroscience, from studies of the brain's molecular and cellular workings, to brain mapping, and all the way up to cognitive psychology.

That, they say, would open the way to some powerful cross-fertilization. "Molecular biologists working on the brain don't read too much cognitive psychology," says Michael Huerta, a neuroscientist who is chief of the neural systems program at NIMH. "But the point of these networks is to allow the integration of this kind of information." Cognitive psychologists could tie their theories about the workings of the brain more closely to actual structure, while molecular biologists could begin to see how the structure in the neurons relates to overall function. Researchers might be able to pinpoint more precisely the underlying brain damage that causes Alzheimer's disease, schizophrenia, or stroke, while those who compare animal brains with humans might find better animal models for studying disease or testing new drugs.

Standing in the way of that vision are some daunting technological and sociological challenges, as well as funding worries. But if brain science's effort at grand unification succeeds, it will provide a model for other fields, such as ecology, that also could profit by merging data from many different lines of investigation.

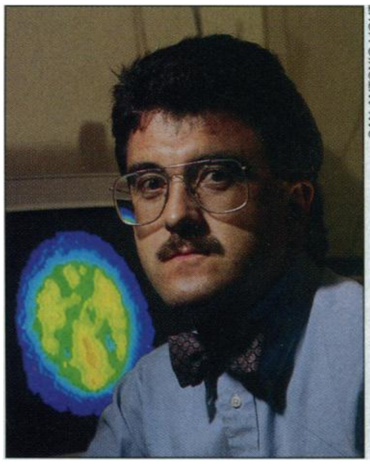
A PET project

The first hints of that kind of unification are already evident in efforts like BrainMap, which is slowly expanding from a single, specialized data catalogue into something more comprehensive. When it all started in the mid-1980s, says University of Texas neuroscientist Peter Fox (BrainMap's main developer along with University of Texas physicist Jack Lancaster), he was just trying to cope with his own PET data. Recalls Fox, then a researcher in brain imaging pioneer Marcus Raichle's lab at Washington University in St. Louis, "It became obvious to me I couldn't keep all the observations of my data in my head," says Fox. "So, I played around with ways to create a database to allow me to explore my data to see if a hypothesis about it was viable or not." Then he heard about the community databases in which geneticists were storing the sequences for genes as soon as they submitted them for publication. Fox realized how useful it would be if neuroscientists also could pool their data, and he convinced several foundations and government agencies to fund his efforts to design a database of brain mapping data.

The result is the "prototype" database that Fox tested on Petersen and three dozen other neuroscientists at the workshop in Texas earlier this month. In its current form, BrainMap contains the data from 30 published journal articles representing several hundred experiments, most of them involving PET data. After BrainMap goes on line with 150 articles—over the Internet in about 6 months—Fox hopes it will grow rapidly. And while BrainMap now comprises primarily PET scanning data, he hopes to expand it to contain images of the brain from other mapping methods, such as magnetic resonance imaging (MRI), magnetoencephalography (MEG), and electroencephalography (EEG), as well as studies of the effects of brain lesions.

*"BrainMap Workshop I," 30 November to 2 December, at the University of Texas Health Science Center in San Antonio.

Harvard University neuroscientist Marsel Mesulam, for one, is looking forward to that broadening: "A criticism of brain lesion studies is that they only tell you what the brain does without the lesioned area, rather than telling you what the lesioned area does," he says. Comparing brain lesion studies with functional brain mapping—in which neuroscientists link motor, visual, and other functions to specific areas—should sharpen the overall picture of the brain. Neuroscientist David Van Essen of Washington University foresees even greater gains if the database is expanded to include PET scans of animal brains, some parts of which are known in far more detail than the human brain because studies can be more extensive. If the database makes it possible to transfer to humans the detailed picture he and other investigators have developed of the visual system in macaques, it could lead to a better understanding of human vision and, perhaps, new therapies for some visual disorders.



SAN ANTONIO LIGHT

BrainMapper. Database designer Peter Fox and his creation.

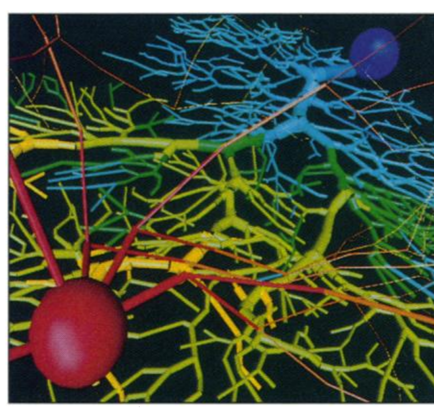
Elsewhere in neuroscience, other databases are springing up that may eventually join BrainMap in the larger network of databases. Using cutting-edge computing technology, Bower at Caltech is turning his computer simulations of brain circuits into a database to try to understand the functional organization of the nervous system. Meanwhile, Scripps Research Institute neuroscientists Floyd Bloom and Warren Young have developed an on-line atlas of the rat brain, known as the Brain Browser, and are trying to expand it. Scripps has hired three software programmers to develop the Browser into something more like an on-line encyclopedia for the community, which eventually should include images and data on the biochemistry and circuitry of the rat brain.

Learning to share

Even before they are linked in a set of community-wide databases, though, these efforts have encountered some of the hurdles that stand in the way of that larger confederation. As BrainMap expands to encompass other kinds of imaging data, for example, it is running into a problem Fox has managed to avoid so far—the need for consistent data standards. The neuroscientists who use PET had already taken care of that problem by setting an international standard for labeling parts of the brain with coordinates rather than names, making it possible to compare data obtained from different brains or by different techniques. But that kind of standardization doesn't come easy. "A lot of fields haven't gotten past this point of how to represent their data," says Bruce Schatz, an information specialist at the University of Arizona, and architect of the Worm Community System (WCS) database, which links researchers studying the nematode worm *Caenorhabditis elegans*.

Bower, in his brain circuit database, has been grappling with another issue that plagues all databases: quality control. "No one wants a database of junk," as he puts it succinctly. By junk, Bower means not just data entered incorrectly (he's got safeguards for those) but also data that are just plain uninteresting. "Just listing the 85 potassium channels is not particularly satisfying if you're interested in how the cell system works," he says. The key, says Schatz, is to establish a quality-control process within the system to make sure the entries are credible and useful.

Fox, meanwhile, has had an early taste of what is bound to be a controversy about what kind of data to archive—and how to control access to it. He'd like to see BrainMap come to include both published and "raw," unpublished data—or at least the complete data sets that were published in condensed form in journals. By giving researchers a more complete picture of their field than published data alone can offer, says Fox, "BrainMap should help labs cut duplication of experiments, and make them more efficient."



JIM BOWER, JASON LEIGH, AND ERIK DESCHUTTER

Circuit city. Genesis models brain wiring.

But the prospect of making raw data available to a wider community raises sticky questions, says Schatz. Databases will need mechanisms not found in commercial software to allow researchers to determine who will see the raw data—whether they limit access to their own lab or a few collaborators, or offer it to the entire neuroscience community. There's also the question of when raw data should be unveiled. Fox raises the possibility of an arrangement with journals publishing in the same field, in which scientists would be encouraged to archive their underlying data in the database when they publish, much as geneticists already do.

Even if Fox, Bower, and other database pioneers get their databases up and running, the ultimate challenge will be merging them. Designers will have to build a single software environment to link lots of labs running separate programs and databases on all kinds of different computers. The object is to allow scientists—even those without special computer skills—to browse easily from database to database. What's needed, says Schatz, is "a system that lets you concentrate on navigating between different pieces of data rather than dealing with each individual database and program."

All this, of course, won't come fast or cheap. The final hurdle will be funding—and estimates of the ultimate cost range from tens of millions to billions of dollars. Beyond the first few million, expected to come early next year from the many federal agencies participating in the Human Brain Project, prospects are uncertain.

Other communities are intently watching neuroscientists' efforts to negotiate these hurdles. Says Dan Sulzbach, executive director of the San Diego Supercomputer Center, who is working with ecologists to build a community database that could include satellite images: "We're interested in how neuroscientists solve the problems of how to store the data, how to access it, and what kinds of queries you use to access images, for example." Ecologists, like archeologists and other researchers, recognize that they too could benefit by making data from many different subfields available at the same time to a single investigator—letting him or her see all parts of the elephant at once.

And there's another impetus for databasing that is certain to grow in the future, bringing more and more fields into the fold: "There will be masses and masses and masses of data," says Sulzbach. "There's going to be so much data that individual researchers won't have the facilities or personnel or interest to manage all that data"—whether it is images of the brain, Earth from space, or the genes in different organisms. Smart databases that can scan this huge universe of data and focus on an object of interest are fast becoming critical tools for future scientists. And as a few scientists master those new tools—and publish the resulting breakthroughs—it will only be a matter of time before their colleagues will want to go on line as well. Says Young at Scripps: "They'll have to move in this direction, or they'll be out of it."

–Ann Gibbons