The Neural Network Is Advancing

By Greg Goth

The potential establishment of a worldwide neuroimaging resource network holds a tremendous amount of promise for researchers and clinicians.

"This field we're in, this so-called 'brain mapping,' is quite unique in science," says Alan Evans, director of the Montreal Consortium for Brain Imaging. "Everybody in the field adheres to the same coordinate space for the brain. When we talk about *xyz* coordinate for some part of the brain, we're all talking about the same point. It's trivial, a simple concept, but it has incredibly powerful ramifications. It means any lab, anywhere, can reproduce the results of another lab in a specific, quantifiable way."

Yet, while the potential for a global neuroscience network is vast, the technological and cultural obstacles to establishing the network are also daunting. As the science of brain mapping developed, it was hamstrung by insufficient processing power and data transmission capabilities to allow widespread sharing of brain imagery and related data. The Human Brain Project, initiated in 1993 by the National Institutes of Health, is in the vanguard of breaking those obstacles down.

In fact, Stephen Koslow, director of neuroinformatics at the National Institute of Mental Health, which is coordinating the Human Brain Project, says the effort to build a workable network has barely begun.

"On a scale of one to 10, I'd say we are probably at two," Koslow says. "There are reasonable efforts here in the US with the Human Brain Project to create databases, and a number of ongoing activities, but not all fields are covered, and we haven't started to get data from people putting it in voluntarily. There are also efforts globally to make this work, and they're in a similar state. The notion is in its infancy, really. It's not dissimilar with what happened with the genome and proteome projects when they first started to share data. There's great resistance, and then, slowly, people start to add data and the whole thing changes."

Koslow says the project has issued about US$20 million in grants to 30 groups working on the technological aspects of building accessible neuroscientific databases and integration tools. Some of them are outgrowths of earlier projects that have been several years in the works, while others are new modifications of existing technologies.

Database necessities

Scientists at the UCLA Laboratory of Neuro Imaging (LONI) have been at the forefront of collecting images of the human brain and in trying to make that collection a resource for scientists far removed from the UCLA campus.

"We recognized that the maps we had of the human brain before we started this project were really quite poor," says Arthur Toga, director of LONI. The problem, he says, was that the lab had very few images of brains, which made mapping a representative brain for any given population extremely problematic. Under the aegis of researchers in the International Consortium for Brain Mapping, which receives contributions from researchers from several sites worldwide, LONI has amassed 7,000 data sets.

"So we embarked on a project of creating, instead of a static representation of the human brain, a probabilistic and ultimately dynamic representation," Toga says. "We wanted to have statistics that would describe the probabilities of having a given structure belonging at a particular coordinate, by including variabilities within the atlas."

"We've been working on this for seven or eight years and have built a system that has some fairly complex mathematics that support it, that lets us calculate these variables, and allows us to create visualizations that represent brains of populations. The problem was, we didn't have a database, and we were collecting data at an alarming rate from centers around the world."

To build this database, the UCLA and ICBM scientists have partnered with MITRE researchers to develop a digital library and data warehouse architecture that can calculate the complexities of three- and four-dimensional queries.

To understand the problems facing the researchers, you must first understand the concept of the large probabilistic atlas, says Ken Smith, MITRE's principal investigator.

"There's quite a bit of variation even among 'normal' individuals, when you see all the folds of the brain," Smith says. "They differ even between monozygotic twins. As you combine more and more of these images, such as all brain scans

1521-9615/02/$17.00 © 2002 IEEE

for Asian males aged 40 to 50, the noise begins to be eliminated, and you begin to see features that are characteristic of that population but not until you can do that over a larger *n*. That's called atlasing."

To house these atlases, the LONI scientists have employed a 100-terabyte StorageTek Powderhorn storage system and Silicon Graphics Clustered XFS file system. The Clustered XFS system, which replaced a network file system architecture, lets users access data from multiple work sites simultaneously via Fibre Channel storage area network connections, eliminating a central file server—a single point of failure—as would be present in a LAN. User interface is Web-based Java, with SQL databases on the back end.

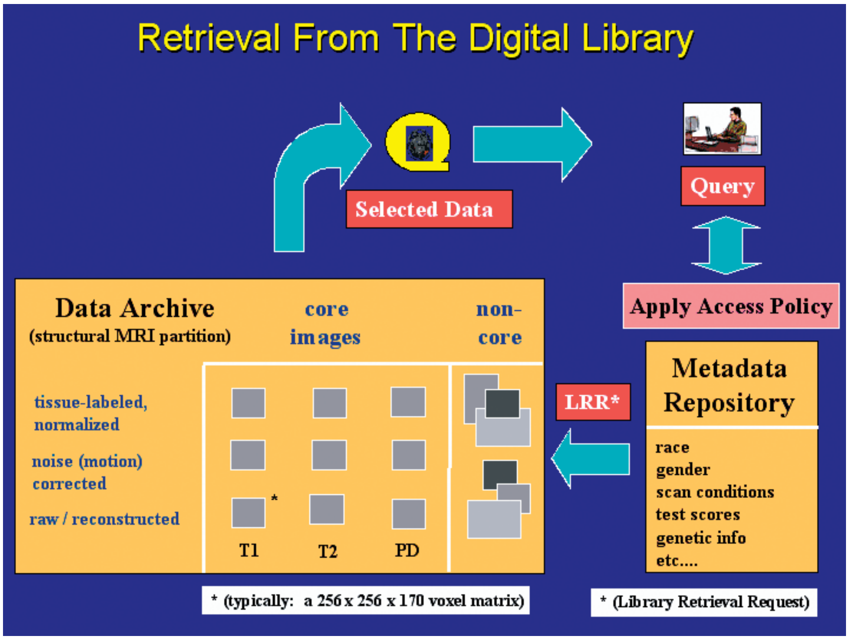


Figure 1. A demonstration of retrieving data from the neuroscientific digital library being created by MITRE researchers.

"The data can be stored either on the Powderhorn or some buffer farther in," Toga says. "There's no single point of failure anywhere."

Smith is six months into a five-year grant to create a digital library of the LONI images and a data warehousing scheme. His initial goal is to build a functioning retrieval system for the library. Smith hopes the library, for both retrieval requests and submissions, will be functional by the project's third year.

"The digital library will simply be for sharing data," he says, "to allow the upload and download of images. People would be able to come to a form and query the system based on a set of attributes of the image itself, such as the parameters under which it was scanned, age of subject, and so on."

Figure 1 shows the step-by-step process of the retrieval method.

Users will interact with a metadata repository maintained in a relational database management system via a Web browser programmed with Java servlets. Through this interface, users can examine the metadata to visualize the database contents (such as an age histogram) and use the metadata values to select imagery (such as brain volumes).

During this process, the software applies an access policy depending on the access restrictions on the data. For example, if a user requests several volumes that require the owner's approval before the user can retrieve them, a contact would prompt the user to request approval.

Once users select the images and the access policy is applied, the system sends a library retrieval request to the LONI data archive. The software then retrieves the images from the silo and sends them to the user.

Building data warehouse and content-based query capabilities will be more complex, because pinpointing specific regions or characteristics of the brain will demand extending current database capabilities.

"Commercial databases offer content-based retrieval of images, and these are useful for photographs of scenery and people," Smith says. "However, the underlying features these databases use (such as color histograms) are insufficient for the special problem of neuroimagery. Many hundreds, or thousands, of domain-specific features would need to be extracted and stored to support content-based retrieval in neuroimages. The selection of these features is a research problem.

"A related problem is that the ability of databases to index high-dimensional feature spaces drops off around 20 features. Thus, efficient search through a large database for similar neuroimages is also an important research problem."

In Montreal, Evans is coordinating the development of a pediatric neuroimage database project comprised of seven sites, each of which will send data from between 50 and 100 children. Presently, researchers at these seven centers can access the project's data (which is stored in a MySQL database) via a standardized laptop interface. Eventually, Evans says, the data will be available to researchers and clinicians at large, which should let them compare characteristics of groups they are studying with the normative database in Montreal.

Sharing tools, not data

Other researchers are working on methods to help neuroscientists integrate neuroimaging analysis tools developed by disparate labs. Because these tool developers have been working on different problems, potential users of these tools

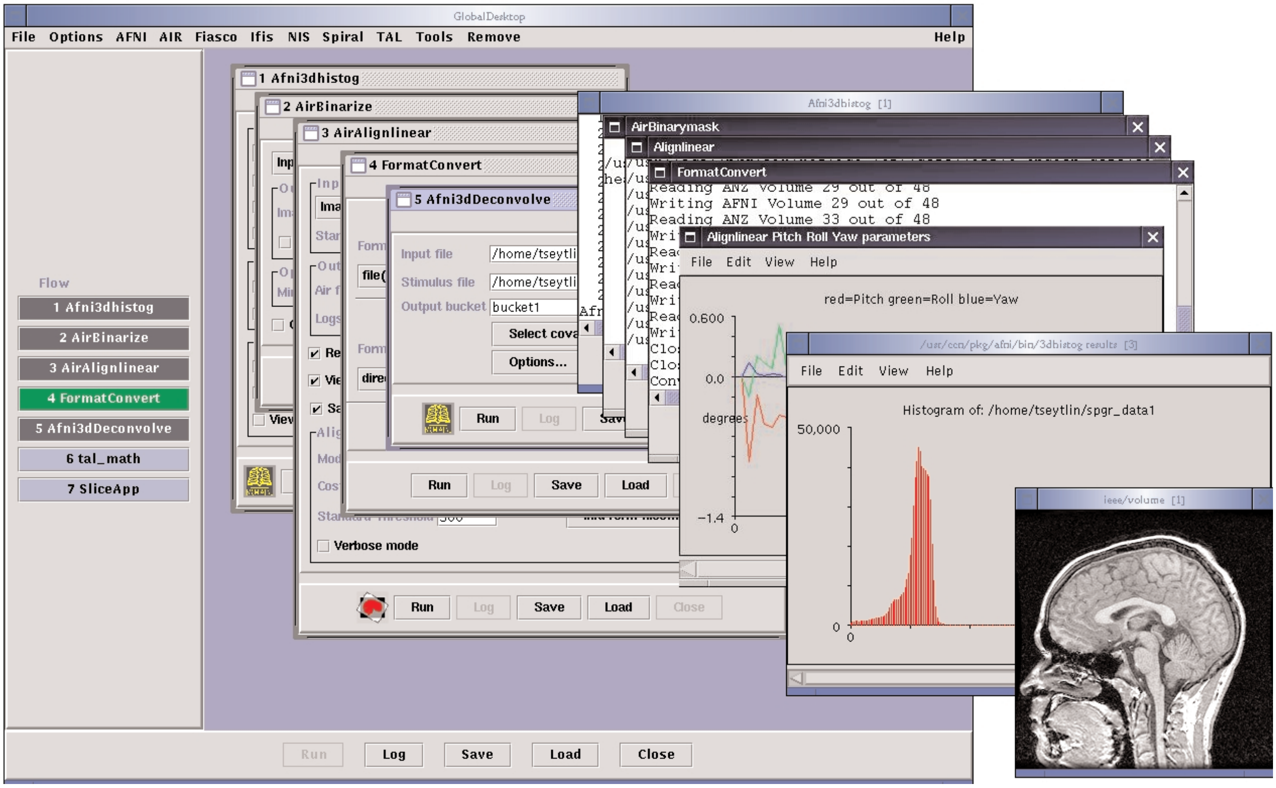


Figure 2. A Fiswidgets desktop running a batch mode processing stream to perform motion correction and multiple regression on brain images. It invokes wrapped applications from the AIR, AFNI, and tal packages.

often must contend with incompatible interfaces, data formats, and computing environments.

Researchers at Princeton University and the University of Pittsburgh are developing a common graphical user interface called Fiswidgets (Functional Imaging Software Widgets) to integrate these various tools (see Figure 2). Fiswidgets is a set of Java classes built on top of Java Swing classes. It is designed to permit programmers to write simple Java graphical user interfaces to run applications that would otherwise be invoked by a command line or from within a shell script. Fiswidgets developers are also working on a graphical analysis desktop to link together individual GUIs as subcomponents in multistep processing streams, and utilities such as medical image format converters and data display tools.

"There are a fairly large number of scientific labs working on imaging problems that do in fact develop their own software with quite high-quality, sophisticated algorithms," says Kate Fissell, Fiswidgets project manager. "These are available to the general public, and we can use them as the scientific components in the Fiswidgets architecture. One of the reasons they're available is that there's really a need for these labs to disseminate their tools. Other researchers want to use these tools, but don't have the programming and computational resources to either write their own programs or to deal with the interoperability problems associated with just grabbing somebody else's code and saying, 'Let's go with it.' "

Fissell says the Fiswidgets wrappers might simplify resource sharing, but she is careful to point out that the project is not exactly a data sharing tool.

"There is a distinction between sharing data and sharing tools. One might think they go hand in hand, but that's not the case, and what we're focusing on is sharing the tools. Sharing data brings up another whole set of issues and reasons about why people share it. Sharing tools is really more of a day-to-day, up-to-your-elbows situation. If you're sharing data, maybe you just need to convert it once to your local format and run with it. However, if you're using the tools from five or six labs in the same processing stream, you might need to juggle things back and forth five times a day."

The Fiswidgets team is developing wrappers for two of the most commonly used neuroimaging tools: Automated Image Registration and Analysis of Functional NeuroImages. AIR, written by UCLA researcher Roger Woods, essentially aligns, or registers, images taken at different times of the same subject, or of different subjects, or via different modalities (such as PET versus MRI); AFNI, a Unix-based software developed by Robert W. Cox at the Medical College of Wisconsin, analyzes and visualizes functional magnetic resonance neuroimages.

"Basically, we're leveraging off of Unix command line programs," Fissell says. "If there's a program with a strong, clean command line, you can put the GUI wrapper on top of that."

Smith sees potential for Fiswidgets in the larger context of the integration effort. "If someone just downloaded an image from our database that was fully unprocessed, as it came off the scanner, they would need to register it using AIR. Using Fiswidgets, they could register any number of images that come from our database. The advantage is that other tools could be developed and used as well, using Fiswidgets. Data is the fuel that comes into these tools."

Fissell says several other teams are working on Fiswidgets-like interfaces, but she does not consider them competitors as much as alternatives. One such effort is the LONI Pipeline, undertaken by the UCLA team, which Fissell says is more focused on linking supercomputer resources than smaller desktop environments; the other is VoxBo, a project developed at the University of Pennsylvania by Dan Kimberg and Geoff Aguirre.

"There's definitely a thought out there that this kind of integration is useful, needed, and possible at this point," Fissell says.

Cultural restructuring is needed

As the technology making distributive neuroscience progresses, the research community must also face the delicate task of devising new methods to reward those who share raw research.

The NIH recently posted a new policy encouraging researchers to share data, but that request is not being met with open arms.

"This issue is extremely sensitive," Montreal consortium director Evans says. "People can potentially see their entire careers disappearing out the door if they're generous. Until the system of academic merit and promotion and excellence recognizes these principles in a tangible way, people are going to be nervous.

"There's money on this. The link between this field and the pharmaceutical industry, the biotechnology industry, the telecommunications industry, the computer industry, and scanner manufacturers is so intense, you could develop a technique and find its commercial expression quite quickly. In theory, you could make a lot of money in a way that's not possible in the arts or other sciences. The NIH can say all it wants, but it has to back up those statements with concrete indices of merit. Someone has to be able to say that in addi-

Useful URLs

Fiswidgets home page:
http://neurocog.lrdc.pitt.edu/fiswidgets

Human Brain Project home page:
www.nimh.nih.gov/neuroinformatics/index.cfm

LONI home page:
www.loni.ucla.edu

MITRE's neuroinformatics home page:
www.mitre.org/neuroinformatics/index.html

NIH data sharing policy statement:
http://grants1.nih.gov/grants/policy/data_sharing/index.htm

tion to first-author publication, if you built a database or contributed a certain amount of data to it, it counts in some tangible way toward academic advancement. Without that, without intellectual ownership of data as well as ideas, people will balk, and it's naive to think they wouldn't."

NIMH's Koslow agrees with Evans that change is needed in the reward system. He believes the NIH's new policy could gradually help change the prevailing cultural mindset. "Actually, on grants now, one of the review criteria is what individuals propose to do in terms of sharing data," he says.

The issue of setting technical standards for neuroscience also looms. Evans says the Organization for Human Brain Mapping, a larger umbrella group of researchers, might spin off a subcommittee to address the issue.

"But that's a dangerous game, because people get very protective of their own way of doing things. When you make recommendations as a committee, you're often standing there to be shot at by various factions. The OHBM may have a stronger credibility to recommend international standards, but the issue is fraught with politics."

LONI's Toga says the prospect of formal standards in neuroscience technology will be a nonstarter. "I think you'll see folks who will build things that will become remarkably well accepted, and then, because of their large constituency, will emerge as standards," he says, noting the AIR program as an example. "A top-down approach in this field will never happen."

Despite the potential obstacles to building a worldwide neuroscience network, Evans says the computational advances have profound implications for researchers and clinicians.

"It's such an exciting time to be in this business," he says. "The speed of the computer allows us to conceive of things we couldn't even conceive of. It's not just a matter of doing things faster now that we could do before.

"If we look at the area of functional connectivity, for example—if *A* fires off, what other parts of the brain fire—what's the temporal ordering between steps as you go from one part of the brain to another? If you think about that, you're looking at a 3D field, plus time, which becomes a 4D data set. For every voxel, you have to look up the time in all the others, so for every 3D voxel, you have to study another 4D data set. The number crunching involved will bring you to your knees. But the tools we have allow us to look at this—how the normal brain works, how the disordered brain works, how the signal propagation from one part of the brain breaks down in schizophrenia or in Alzheimer's disease. All these things become new ways to think about studying the brain, which are 100 percent quantitative and mathematically reproducible, but have direct brain research and clinical neurological consequences. The field has moved away from description and finger-pointing."

Greg Goth is a technology writer based in Oakville, Connecticut. Contact him at gregorygoth@compuserve.com.

Nanotechnology Meets Marine Biology

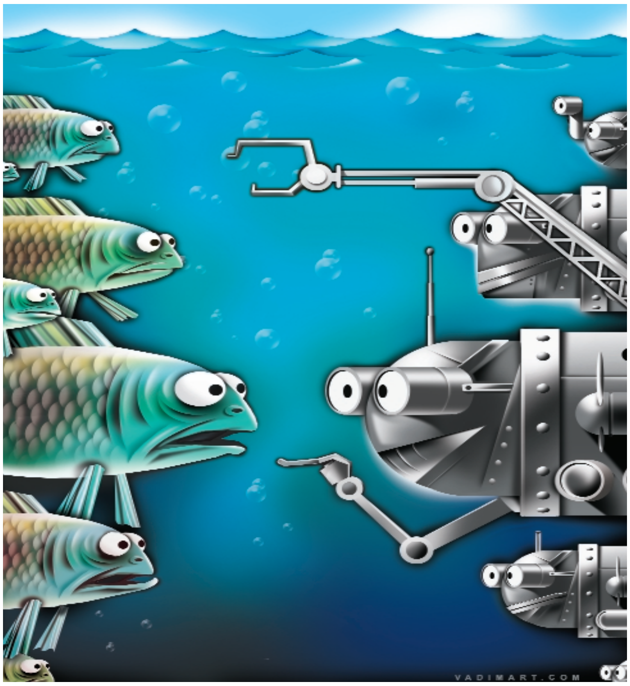
By Anne Jacobson

A decade from now, some of the tiniest and most unusual inhabitants of the ocean might be schools of aquatic nanorobots deployed to monitor the water's health.

Researchers at the University of Southern California School of Engineering recently received a $1.5 million grant from the US National Science Foundation to create millions of sensor-equipped, microscopic robots that would scour the ocean for potentially dangerous microorganisms.

"The oceans can have a lot of nasty things in them," says Ari Requicha, director of USC's Laboratory for Molecular Robotics and lead investigator on the project. "Here in California, beaches close every year because of bacteria. Recently, dolphins and sea lions have been dying in Southern California, apparently because of a toxin secreted by a microorganism. The quicker we can learn that a pathogen is present in the water, the sooner we can warn people and begin action to correct the situation."

Although Requicha estimates at least a decade before the first working nanorobots take to the water, his team has made progress. Working with an atomic force microscope—a tool that can both sense and manipulate the tiniest objects, even down to individual atoms—the researchers have precisely placed nanometer-sized particles such as colloidal gold and silver balls onto tiny slabs of silicon and assembled them into a single-electron transistor and an optical waveguide.



Next, they plan to fashion a nanoscale switch.

Eventually, the team will build bacteria-sized robots that can propel themselves, sense problem-causing microbes, communicate via electronic signals, and perform limited computations. The USC researchers must also create software to coordinate the actions of a million or more robots, which Requicha describes as a daunting task.

"I don't think these robots will be confined to the ocean," he says. "We will eventually make robots to hunt down pathogens or repair cells in the human body."

BIOINFORMATICS BOOMING

By Anne Jacobson

Information technology companies serving the life sciences industry can look forward to a healthy market over the next several years. While growth in other IT markets has lagged due to the failure of many dotcoms, the $12 billion "Bio-IT" market will grow 24 percent annually to almost $38 billion by 2006, according to new research from the International Data Corporation, a global market analysis and consulting firm.

"The difference between this trend and the dotcom bubble that burst in 2000 is that the life sciences industry is established; it has been around for centuries and will continue to exist because medicine and health care are basic human needs," says Mark Hall, IDC research director for life sciences.

The life sciences industry will spend heavily on IT as it increasingly relies on computing power to discover new drugs and to unravel the complexities of the human genome. Pharmaceutical and biotech companies in particular will fuel this trend because they purchase more high-performance computers and servers, more powerful database technologies and tools, and increasingly sophisticated software.

Whereas research biologists once sat at lab benches and tested potential new drugs by hand, that work is increasingly done by automated robotic systems that test thousands or even hundreds of thousands of chemical compounds at once, Hall says. Once generated, scientists then sort through data with algorithms that help identify good drug candidates.

Using computational technologies to identify new genes and their functions, biologists will likely mine the data in the human genome sequence for years to come, thereby sustaining the unprecedented growth of the Bio-IT market, Hall adds.

As databases grow, storage will be the life sciences industry's chief need. By 2006, research organizations will spend $11.8 billion on storage and data management systems for the huge amounts of information they generate. After storage, the industry will need better servers to handle increasingly complex workloads. Spending on servers will likely reach $10 billion by 2006, the IDC study concludes.