PAGE 2253
Ice Man thawed for science

2255
The size of Gemini's kneecap

DATA NETWORKS

Scientists Weave New-Style Webs To Tame the Information Glut

Physicists collaborating on a new generation of big experiments may drown in a data waterfall unless they find a way to channel the flow. A consortium of 16 universities has just received an $11.9 million federal grant to build a shared computational network, or data grid, that they hope will serve as the right sort of pipeline—and lead to even better science.

The idea behind data grids is to allow users to tap into a universe of electronic information, regardless of its location or origin. The grids are often compared to the popular music file sharing program Napster, which enables Internet surfers to exchange files. But Napster still relies on a central server to keep track of which music clip is on whose PC. A better comparison is a rival program called Gnutella, which allows users to share any file format in a totally decentralized system.

University researchers want to do the same trick with supercomputers and large data sets. To do so they've created a consortium, funded in part by the National Science Foundation, called the Grid Physics Network or GriPhyN (pronounced "griffin"). "GriPhyN will solve problems more demanding than any individual can solve," says Ian Foster, a computer scientist at Argonne National Laboratory in Illinois and co–principal investigator of the GriPhyN project. Biologists and medical researchers have also seen the value of peer-to-peer networking (see below) and want to make their data available over grids, too.

Right now, physicists can share big databases, but it is a nightmarish task. "We've been doing this for a long time, but it requires a lot of special expertise," says Fabrizio Gagliardi, a CERN physicist heading DataGrid, a European project that will join with GriPhyN. "Right now you have to know the exact locations and access procedures for each computer system." He compares it to e-mail 15 years ago: "When I was working at Stanford, I had to log in to five different machines just to read my mail at CERN." Data grids will make global data sharing painless, Gagliardi says.

GriPhyN is arriving just in time to serve several large physics projects. Initially it will join the Sloan Digital Sky Survey (SDSS), the Laser Interferometer Gravitational Observatory, and two experiments at CERN, called ATLAS and CMS, that will run on the Large Hadron Collider (LHC). Each project offers the type of challenge that GriPhyN hopes to conquer: oceans of data that thousands of collaborators around the world must analyze to pick out painfully small signals from a noisy and cluttered background.

EXPERIMENTS SERVED BY GRIPHYN PROJECT

Application	Data volume (terabytes/year)	Type of data
SDSS, 1999	10	Catalogs, image files
LIGO, 2002	250	Multiple channel time series, Fourier transformations
ATLAS/CMS, 2005	5000	Events, 100 Gb/sec simultaneous access

When the LHC comes online in 2005, for example, the collisions of its subatomic particles will generate a data stream of 5 petabytes every year. One petabyte is roughly equivalent to the capacity of a million

Downloading the Human Brain, With Security

Neuroscientists collect huge quantities of data on the human brain. But compared with their colleagues in physics, they are traditionally much less likely—for professional and personal reasons—to want to share them (*Science*, 1 September, p. 1458). Now a group at Rutgers University in Newark, New Jersey, is proposing a two-step, encrypted process for sharing information that would open the door to all legitimate researchers while imposing tough safeguards on its use.

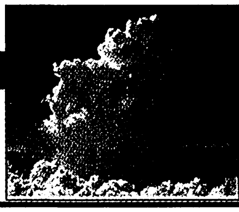
The Rutgers group is writing software for a Napster-like Web site that would make possible "peer-to-peer" sharing of brain-scan images, but would not contain images itself. Instead, the site would house an index of available data sets and a protocol for accessing them. Researchers willing to share brain images would register with the site and describe what was available and under what conditions, says Benjamin Martin Bly, a cognitive neuroscientist at Rutgers who, along with his boss, cognitive scientist Stephen Hanson, is developing the project. "You might look up language localization [in the brain]," Bly explains. "The server would respond with a list of 100 results," each describing an experiment and the data, along with the conditions on access imposed by the donor. If the seeker agreed to the conditions, the central server would forward a request to the donor. If the donor also agreed, submitting an encrypted signal, the server would automatically trigger a "handshake" that would download the file.

Bly says the system will "separate the information about what exists—which can be shared easily—from the actual experimental data." The operators of the central site would never handle the raw data. This kind of protection, he adds, would reduce concerns about breach of confidentiality—always an issue in clinical studies—and increase donor confidence.

The protocol would differ from Napster's, Bly points out, by letting donors control the release of data and keeping a record of each handshake. If a person claims to have data of a certain type but refuses to share it, "this protocol makes it immediately obvious." Bly is designing the software with encouragement from the National Institute of Mental Health and hopes to have a test or "beta" version ready by January.

—ELIOT MARSHALL

CREDITS: (TOP TO BOTTOM) GRIPHYN; R. C. COGHILL, J. G. MCHAFFIE, AND Y. YEN/WAKE FOREST UNIVERSITY SCHOOL OF MEDICINE



personal computer hard drives or a stack of CD-ROMs nearly 2 kilometers high. Particle physicists have always had to deal with reams of data flooding from their detectors, but nothing like LHC. "The current experiments at CERN's Large Electron-Positron collider generate a few terabytes per year," says David Strickland, a Princeton physicist working at CERN. "The new experiments will create 1000 times more data than that"—data that thousands of collaborators will need to find and use.

The researchers also need massive computer muscle to crunch the numbers. In SDSS, for example, an astronomer trying to puzzle through a possible case of gravitational lensing might need to sort through 10 million galactic objects in order to find an effect, using sophisticated statistical wizardry and careful mathematical filtering. "The resources required, for economic or political reasons, just cannot be created at any single location," says Foster.

Physicists now spend heaps of time locating data files and getting them processed. In searching for images representing how new particles created in a collision were slamming around the detector chamber, for example, a researcher may have to punch in a horrendous chain of computer commands to translate the raw numbers into a useful picture of the results.

GriPhyN and DataGrid will also work together with the Particle Physics Data Grid (PPDG) at the California Institute of Technology (Caltech) in Pasadena to provide a connected computational lattice for big physics experiments. Harvey Newman, a physicist at Caltech, is a principal researcher on PPDG and also a senior member of the GriPhyN group. "We'd always planned to study data transfer and file caching in the short term, then build a longer life system," Newman says. "Then GriPhyN came along." Farther afield, the European Commission's Information Society Technologies program has just invited DataGrid to apply for EU9.8 million (about US$8.6 million) to build research grids in Europe.

The international culture of physics fosters such grand sharing, but pitfalls may loom. "We're all nervous about this," says physicist Paul Avery, co-PI of the project at the University of Florida, Gainesville. "My experience in large software projects is that unless you sit on top of this all the time, you do diverge." Newman was also nagged by doubts initially after hearing people discuss the idea at a meeting. "Some people said, 'We'll build our national grids and then make them work together.' But this does not work," Newman says. "Fortunately we haven't built anything yet, so there is a good chance that we'll all build the same thing."

Strickland, who is not directly involved in the grid construction, says that the grid builders appear to be taking the right tack by funding software engineering rather than just buying lots of new hardware: "They seem to be throwing the right resources at the problem." But that alone is no guarantee of success, he cautions. "Obviously, we've got a long way to go." –DAVID VOSS

ANTHROPOLOGY

Misconduct Alleged in Yanomamo Studies

E-mail has been ricocheting among anthropologists as they nervously await the publication of a new book that charges some prominent researchers with professional misconduct—and much worse—in their studies of the Yanomamo, a native people in the Brazilian and Venezuelan Amazon. Written by journalist Patrick Tierney, *Darkness in El Dorado* (W. W. Norton)—an excerpt of which is scheduled for publication in *The New Yorker* next month—accuses anthropologists of creating a false picture of the Yanomami, manufacturing evidence, and perhaps setting off a fatal measles epidemic. "This is the Watergate of anthropology," says Leslie Sponsel of the University of Hawaii, Manoa. "If even some of the charges are true, it will be the biggest scandal ever to hit the field."

Although few anthropologists have actually read the book, which will not be published until mid-November, it has already stimulated an enormous reaction. The American Anthropological Association (AAA) has promised to hold a session on the book at its upcoming annual meeting. Napoleon Chagnon, a prominent Yanomamo specialist now at the University of California, Santa Barbara, whose research is challenged by Tierney, has already refused to participate in what he calls "a feeding frenzy in which I am the bait." (Instead, he is consulting libel lawyers.) Meanwhile, other researchers are recruiting statements to defend the late James V. Neel, a University of Michigan geneticist whom Tierney charges with distributing a measles vaccine in 1968 that may have worsened or even caused an epidemic that led to "hundreds, perhaps thousands" of deaths, say those who have read galleys of the book.

"Yanomamo anthropology has been a battleground for years," says one anthropologist with extensive experience in the area. "But the scale of these allegations is far beyond anything I've ever heard of before." The researcher, who requested anonymity for fear of being drawn into litigation, adds, "The prime rule for anthropology is not to harm the people you're working with. ... This book is apparently saying that researchers have grotesquely violated those standards for 30 years."

The debate over *Darkness in El Dorado* is the latest, biggest skirmish in a decades-long battle over the Yanomamo. Living in more than 200 small villages near the headwaters of the Orinoco River, the 24,000 Yanomami are among the least Europeanized people on Earth. Although missionaries contacted them in the 1950s, the first long-term anthropological study of the Yanomami was not published until 1968, when Chagnon published *Yanomamo: The Fierce People*. Based on his University of Michigan dissertation, the book was quickly acclaimed as a classic, selling almost a million copies and becoming fodder for introductory anthropology courses across the globe. Meanwhile, Chagnon entered into collaborations with Neel, who was beginning a long-term study of Yanomamo genetics, and Tim Asch, a documentary filmmaker. (They eventually made 39 films, several of which

Anthropological warfare. Allegations have reopened a bitter battle over depiction of Yanomami as warlike.

CREDIT: G. PRANCE/VISUALS UNLIMITED