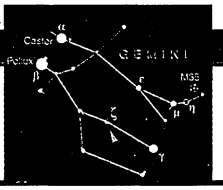
PAGE 2253
Ice Man thawed for science



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.

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

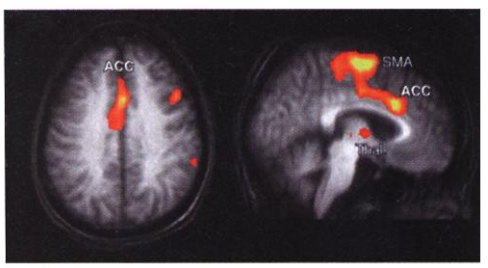
EXPERIMENTS SERVED BY GRIPHYN PROJECT

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

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

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