# SCIENCES
ON THE GRID

Bioscience

Geoscience

Particle Physics

All fields of science benefit from more resources and better collaboration, so it's no surprise that scientific researchers are among the first to explore the potential of grid computing to connect people, tools, and technology. Physics and biology were among the earliest adopters, but chemistry, astronomy, the geosciences, medicine, engineering, and even social and environmental sciences are now kick-starting their own efforts. Here is a small sampling of some of the projects now pushing the limits of grid computing.

**by
Katie Yurkewicz**
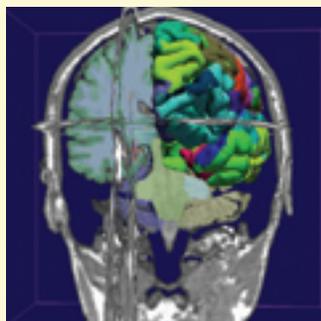
3D Slicer visualization of a brain, obtained from an MRI scan. Colored areas show brain structures automatically detected by FreeSurfer.

Image: Morphometry BIRN

Bioscience

Identifying Alzheimer's disease before a person exhibits symptoms; learning the function of all the genes in the human genome; finding drugs to cure and prevent malaria: From large dedicated biomedical infrastructures to small individual applications, grid computing aids scientists in their quest to solve these and other biological, medical, and health science problems.

One of the first, and largest, of the dedicated cyberinfrastructures is the Biomedical Informatics Research Network (BIRN). Launched in 2001, the National Institutes of Health-funded project encourages collaboration among scientists who traditionally conducted independent investigations. BIRN provides a framework in which researchers pool data, patient populations, visualization tools, as well as analysis and modeling software.

In one of BIRN's three test beds, magnetic-resonance images from small groups across the country are pooled to form a large population for the study of depression, Alzheimer's disease, and cognitive impairment. A large group of subjects makes for a very comprehensive study, but comparing MRI scans taken at different institutions is a challenge worthy of grid computing.

"If you gave several people different cameras from different makers and asked them to take the same picture, they'd all come out a little different," explains Mark Ellisman, director of the BIRN Coordinating Center at the San Diego Supercomputer Center. "If you want to search MRI images for small structural differences in the brains of patients with Alzheimer's disease, you need to make the data from all the MRI machines comparable. This requires methods to align, measure, analyze, and visualize many different types of data, which researchers can develop and share using the BIRN framework."

While BIRN brings together hundreds of scientists in a dedicated infrastructure, a handful of scientists in France are grid-adapting an application to increase the accuracy of cancer treatment.

Medical physicists planning radiation therapy treatment must simulate the passage of ionizing radiation through the body. GATE, the GEANT4 Application for Tomographic Emission, provides a more accurate method for such simulations than many existing programs, but its long running time makes it inefficient for clinical settings. By running in the Enabling Grids for E-sciencE (EGEE) infrastructure, researchers have decreased by a factor of 30 GATE's running time. Medical physicists at the Centre Jean-Perrin hospital in France are now testing the use of GATE and EGEE to increase the accuracy of the treatment of eye tumors.

"Without the resources shared on EGEE, one simulation would take over four hours to run," says graduate student Lydia Maigne, who has adapted GATE to work on the grid. "When split up into 100 jobs and run on the grid, it only takes about eight minutes."

In the field of genomics, the amount of data available has exploded in recent years, making the analysis of genome and protein sequences a prime candidate for the use of grid resources. Computer scientists and biologists at Argonne National Laboratory have developed a bioinformatics application to interpret newly-sequenced genomes. The Genome Analysis and Database Update system (GADU), which runs on the Genome Analysis Research Environment, uses resources on the Open Science Grid and the TeraGrid simultaneously to compare individual protein sequences against all annotated sequences in publicly available databases.

Every few months, researchers run GADU to search the huge protein databases for new additions, and compare the new proteins against all those whose functions are already known. The use of grid computing has greatly reduced the time for a full update, which had skyrocketed due to the exponential increase in the number of sequenced proteins.
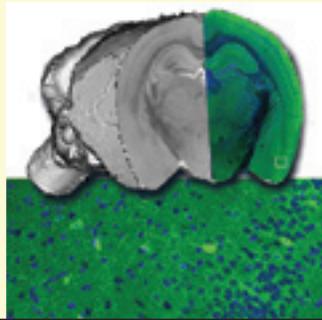
Mouse BIRN researchers are using multi-scale imaging methods to characterize mouse models of human neurological disorders.

Image: Diana Price and Eliezer Masliah, Mouse BIRN

Geoscience
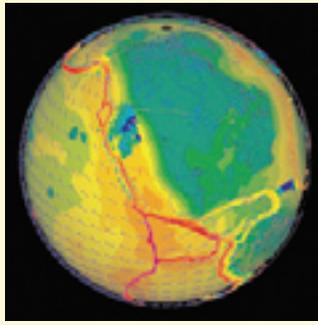
A comprehensive understanding of the Earth's evolution over time, or of the effect of natural disasters on geology and human-made structures, can only be achieved by pooling the knowledge of scientists and engineers from many different sub-specialties. Two US experiments are examples of how the geoscience and earthquake engineering communities are exploring grid computing as a tool to create an environment where sharing of ideas, data, as well as modeling and visualization resources are commonplace.

The researchers of the Geosciences Network (GEON) are driven by their quest to understand quantitatively the evolution of the North American lithosphere, or Earth's crust, over space and time. GEON will integrate, and make accessible to the whole community, data and resources currently accessible to only a few experts. Tools currently under development, like smart search technology, will allow scientists looking for information outside their specialties to find what they need without knowing too much information about a specific data set or database.

"The Earth is one unified system," says geophysicist Dogan Seber, GEON project manager. "To really understand it, we need an infrastructure that any of us can easily use to extract information and resources from other disciplines. My interest is intracontinental mountains, and using geophysics I can understand only some aspects of the system. To understand the whole thing—how they started, why they started, what's happened over time—I also need to know about sedimentation, tectonics, and geologic history."

The George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) seeks to lessen the impact of earthquake-and tsunami-related disasters by providing revolutionary capabilities for earthquake engineering research. Its cyberinfrastructure center, NEESit, uses networking and grid-computing technologies to link researchers with resources and equipment, allowing distributed teams to plan, perform, and publish experiments at 15 NEES laboratories across the country.

NEESit researchers develop telepresence tools that stream and synchronize all types of data, so that key researchers can make informed decisions during an experiment even if they're sitting thousands of miles away. After the experiment, researchers will upload data to a central data repository, access related data, and use portals to simulate, model, and analyze the results.

"Scientists can or will participate remotely in experiments, such as tests of wall-type structures, at the University of Minnesota; long-span bridges, at the University of Buffalo; or the effect of tsunamis using a wave basin, at Oregon State," says Lelli Van Den Einde, assistant director for NEESit Operations. "Our goal is that researchers will upload their data soon after a test has occurred, and once the data have been published they will be made available to the public."

Both GEON and NEES began as US experiments and are funded by the National Science Foundation, but they are making an impact worldwide. Researchers in Korea are developing their own version of NEES, and through GEON, countries in South America and Asia are discovering how to use grid technology to understand the Earth.

Particle Physics
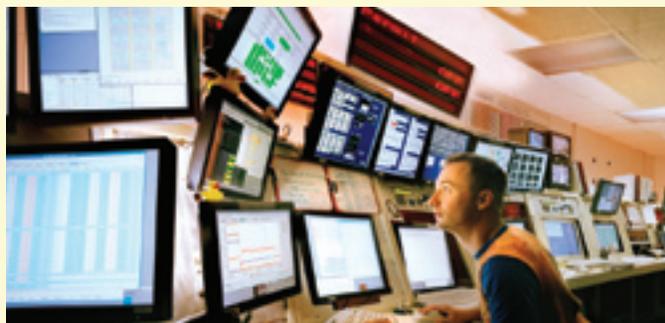
While biomedicine and geoscience use grids to bring together many different sub-disciplines, particle physicists use grid computing to increase computing power and storage resources, and to access and analyze vast amounts of data collected from detectors at the world's most powerful accelerators.

Particle physics detectors generate mountains of data, and even more must be simulated to interpret it. The upcoming experiments at CERN's Large Hadron Collider—ALICE, ATLAS, CMS, and LHCb—will rely on grid computing to distribute data to collaborators around the world, and to provide a unified environment for physicists to analyze data on their laptops, local computer clusters, or grid resources on the other side of the globe.

"When data become available from the ATLAS detector, we will want to sift through it to find the parts relevant to our research area," says Jim Shank, executive project manager for US ATLAS Software and Computing. "The data sets will be too large to fit on users' own computing resources, so they will have to search through and analyze events using grid resources. We are now dedicating a lot of manpower to building a system that will allow physicists to access the data as easily as if they were working at CERN."

While the LHC experiments might be making more headlines recently, Fermilab's DZero experiment has used grid computing to reprocess one billion events over the past six months. DZero has a long history of using computing resources from outside Fermilab, and since 2004 has been using the SAMgrid to create a unified computing environment for reprocessing and simulation.

The reprocessing of stored data is necessary when physicists have made significant advances in understanding a detector and computer scientists have optimized the software to process each collision event more quickly. With the DZero resources at Fermilab busy processing new data constantly streaming in, the collaboration has to use outside resources for the reprocessing.

"The reprocessed data has a better understanding of the DZero detector, which allows us to make more precise measurements of known particles and forces, and to better our searches for new phenomena," explains Fermilab's Amber Boehnlein. "We never would have been able to do this with only Fermilab resources."

With the reprocessing successfully completed, DZero researchers will now use SAMgrid for Monte Carlo simulations of events in their detector. Monte Carlo programs are used to simulate a detector's response to a certain type of particle, such as a top or bottom quark. Without simulated data, physicists could make no measurements or discoveries. Thus, all particle physics experiments generate vast numbers of Monte Carlo simulations. Many—including DZero, ATLAS, CMS, and BaBar at SLAC—have turned to grids for help.

"US CMS has been using grid tools to run our Monte Carlo simulations for over two years," says Lothar Bauerdick, head of US CMS Software and Computing. "It's much easier to run simulations than analysis on the grid, because simulations can be split up into many smaller identical pieces, shipped out across the grid to run, and brought back together in a central place."

Physicists on the LHC experiments, and those looking to build even larger and more data-intensive experiments in the future, hope that the success of the grid for Monte Carlo simulation will be replicated for data analysis, enabling scientists to gain a greater understanding of the universe's fundamental particles.

Top Photo: Control room of Fermilab's DZero experiment.

Photo: Peter Ginter

Left Photo: Magnets installed in the tunnel of CERN's Large Hadron Collider.

Photo: Maximilien Brice