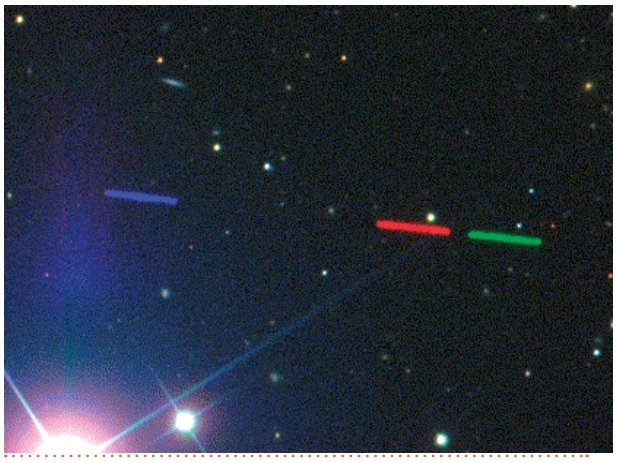
Open Science Grid Applications

The Sloan Digital Sky Survey – The Universe on the OSG

Astronomy on the grid is all about data distribution, and Fermilab's experimental astrophysics group uses all the tactics at its disposal. Three research projects use opportunistic computing and storage resources on the OSG. The first aims to understand the physics of supermassive black holes by fitting a hundred parameter models to 50,000 quasar spectra obtained with the Sloan Digital Sky Survey. The fits require small numbers of files and long compute times, and match well to independent transfers and jobs.



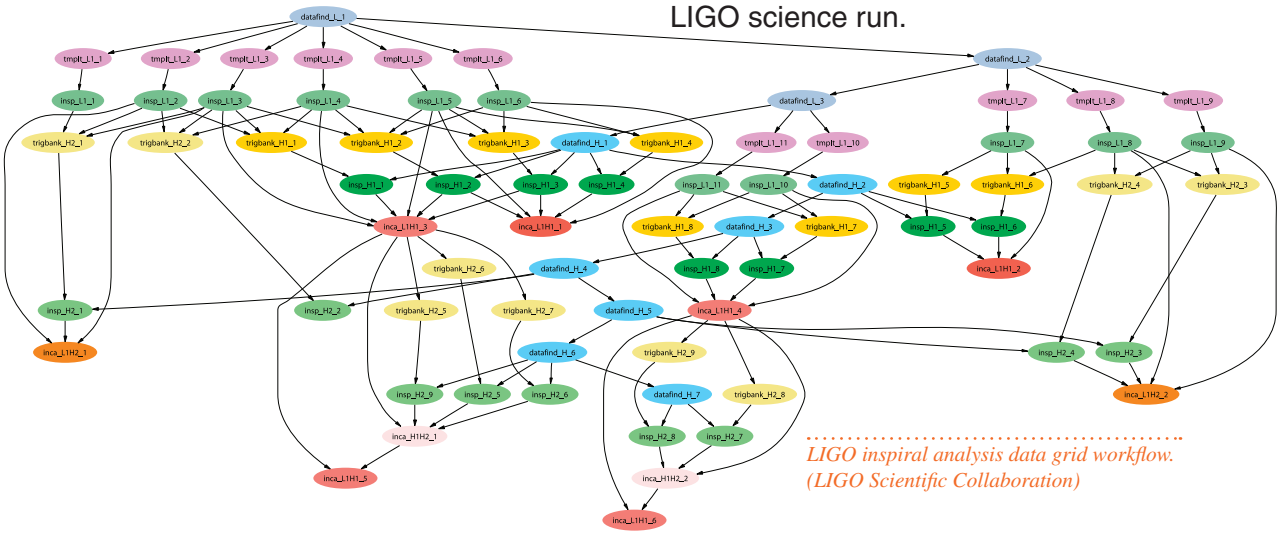
An object passing close to the Earth is captured by the SDSS imaging camera. (Steve Kent)

The second project searches for near Earth asteroids in the SDSS data to understand the statistics of this potentially lethal population. The search involves 500,000 megabyte-scale files which are turned into tarballs and distributed to compute nodes. The third project, the SDSS Southern Coadd, averages repeat observations into a single large and deep image to detect clusters at high redshift. This involves four terabytes of input data, more if independently distributed, and requires pre-positioning of data onto grid storage for local processing.

Searching for Gravitational Waves with the OSG

The Laser Interferometer Gravitational-Wave Observatory is an ambitious NSF-funded project to detect gravitational waves from astrophysical sources such as coalescing neutron stars, black holes, spinning neutron stars and supernovae.

LIGO data analysis is a formidable challenge requiring large amounts of computational power. The coalescing binary searches require approximately 50 million floating-point operations per byte of raw data. Grid computing is an essential component of this search. A workflow used to search for gravitational waves from binary neutron stars and black holes involves analyzing the data from three observatories, looking for coincidences between triggers and performing follow-up analysis for candidate events. LIGO uses the GriPhyN Virtual Data System to plan workflows for execution on the OSG using Condor-G. During SC05, LIGO used the OSG to search for gravitational waves from binary neutron stars in data from the fourth LIGO science run.



LIGO inspiral analysis data grid workflow. (LIGO Scientific Collaboration)

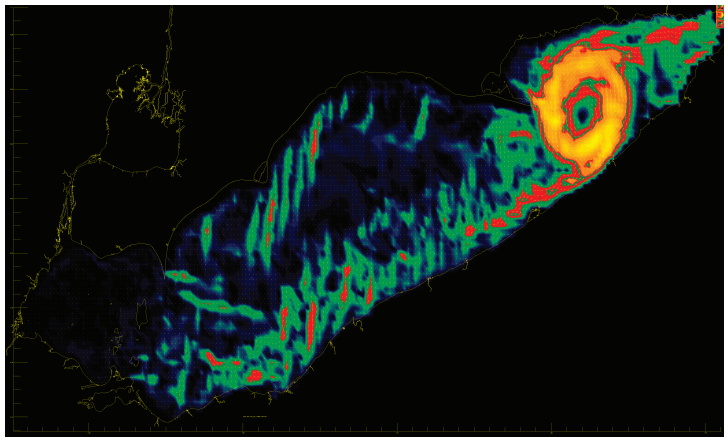
DOE Office of Science

www.opensciencegrid.org

Open Science Grid Applications

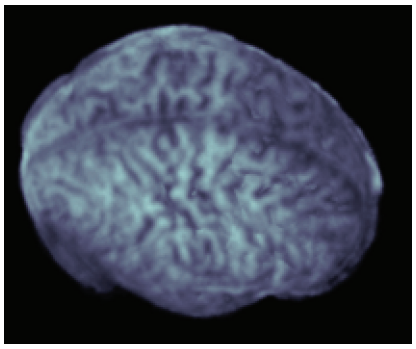
Tracking Algal Blooms in the Great Lakes

Researchers in the Grid Resources for Advanced Science and Engineering (GRASE) VO run hydrodynamic models continuously during the algae growing season to track observed harmful algal blooms within the Great Lakes. The OSG provides computational resources that enable bloom movements to be forecasted for lengths of time ranging from from several days to several weeks. The GRASE VO also provides a science gateway for several other domain science and engineering disciplines, including bioinformatics and earthquake engineering.



Lake Erie surface water velocities. (GRASE VO, Center for Computational Research, University at Buffalo)

Analyzing MRI Data with Grid Resources

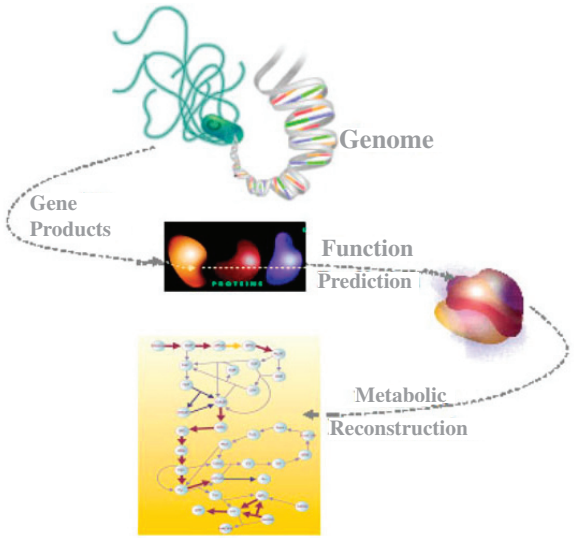


(Scott T. Grafton, Dartmouth Brain Imaging Center and Jed Dobson, Dartmouth College)

Scientists at Dartmouth College use OSG resources to analyze magnetic resonance imaging data. Diverse MRI images collected at the Dartmouth Brain Imaging Center and collaborating institutions worldwide must be normalized before researchers can use them to study the brain. The computationally and data-intensive normalization processes tens of thousands of three-dimensional images for each study. The GriPhyN Virtual Data System runs the large number of data sets in parallel through the normalization workflow. The VDS has been used to abstract some of the workflow pipeline design and execution, and researchers can now run the normalization on local computer clusters, the Dartmouth campus grid or the OSG.

GADU/GNARE – Genome Analysis on the OSG

In the past decade, the scientific community has witnessed an unprecedented accumulation of gene sequence data and data related to the physiology and biochemistry of organisms. A powerful approach to the interpretation of newly sequenced genomes is comparative analysis against all annotated sequences in publicly available resources. GADU pre-computes analysis results for every sequence, finding protein similarities (BLAST), protein family domains (BLOCKS) and structural characteristics. High-throughput computations in GADU are performed using heterogeneous grid computing resources on the OSG, the TeraGrid and the DOE Science Grid. Multi-step genome analysis workflows are expressed and controlled by a "virtual data" model using the GriPhyN Virtual Data System and other grid technologies such as Condor-G and Globus.



High-Throughput Genome Analysis using GADU/GNARE. (Dinanath Sulakhe)

DOE Office of Science

www.opensciencegrid.org

sc05-inserts.indd 4 10/27/2005 7:05:27 PM