Open Science Grid

The Open Science Grid – Its First Year

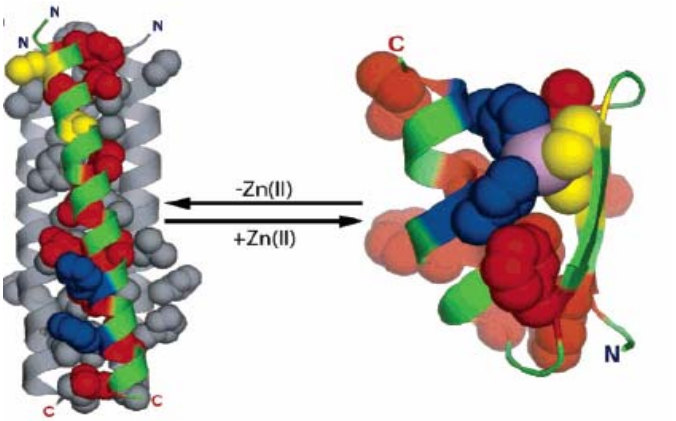
During its first twelve months the Open Science Grid (OSG) Project advanced the state of the art in large scale distributed computing by establishing a novel organizational and operational framework that effectively delivers advanced computing methodologies and capabilities to communities in the life sciences while enhancing and expanding the distributed computing foundation of its founding physics stakeholders. The OSG Consortium has attracted more than ten new institutions to join its shared cyberinfrastructure, further evolved the Virtual Data Toolkit (VDT) as a foundational common middleware distribution, and trained more than one hundred twenty researchers at hands-on grid schools in the US and abroad.

The CERN Large Hadron Collider (LHC) experiment preparations for data taking continue to drive the capability and capacity of the OSG facility. OSG has advanced the information frontier for LHC physics through federating our US distributed facility with partners in Europe and worldwide. OSG is making pioneering contributions in the international arena to ensure global interoperation of data distribution and analysis capabilities for LHC physics and other communities. The close cooperation between OSG and the LHC continues to pave the way for other sciences to benefit from the technology and infrastructure accomplishments.

Through hands-on engagement and training OSG has actively worked to bring the advanced technologies it provides to the nation's university campuses, educators and students, and to bridge the "digital divide" and gender gap. Of the attendees at the training schools to date, 12% are from minority backgrounds and 20% are women. Within the OSG project itself there are three women among the coordinators—the Executive Director, and the Education and Security coordinators. .

Thanks to the open collaboration offered by the OSG Consortium through its participant institutions, middleware providers, dedicated staff and distributed computing resources, a number of research efforts have achieved computational accomplishments unfeasible with conventional computing infrastructure.

Protein design researchers in the Kulhman Laboratory in the Biochemisty and Biophysics department at the University of North Carolina, Chapel Hill, Medical School have successfully designed structures for ten proteins on the OSG. They routinely submit their workloads, of up to a thousand "ad-hoc" jobs consuming opportunistically available resources, onto about a dozen OSG sites.



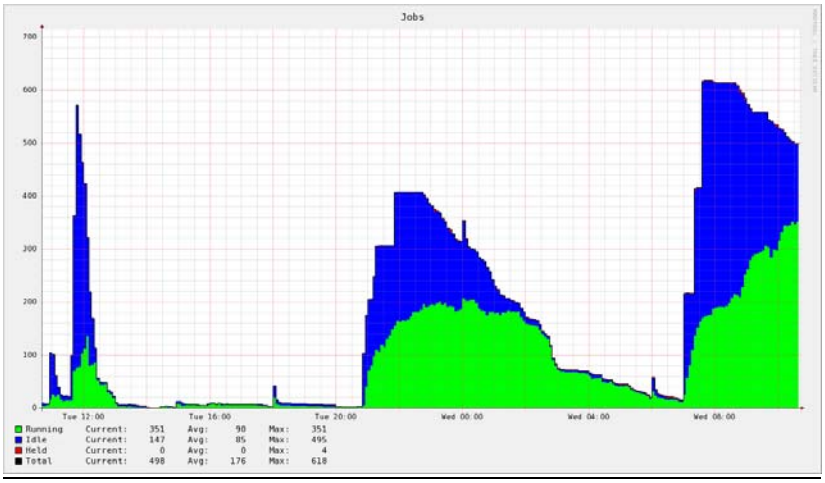


Figure 2: Sample of Periodic Runs of Rosetta on OSG

Figure 1: Computationally designed protein that switches between a zinc finger and coiled-coil depending on environment. Image courtesy Kuhlman Laboratory

OSG has facilitated early accomplishments for a new community of earth scientists using a climate modeling application. This work has already resulted in a joint paper by Brian J. Etherton of University of North Carolina Charlotte and OSG staff member Leesa Brieger, entitled "Probabilistic QPF using a multi-physics WRF ensemble."

In an example of broader benefit facilitated through OSG's collaboration across scientific domains, Ana Damjanovic and her team at Johns Hopkins University are actively using the OSG to study water penetration in staphylococcal nuclease. They run macromolecular simulations using the Chemistry at Harvard Molecular Mechanics (CHARMM) program coupled with the workload management system provided by the LHC U.S. ATLAS collaboration.

"CHARMM is running beautifully on the grid" Ana says. "We should be preparing a scientific publication with the results obtained on the OSG as soon as the final set of runs is over."

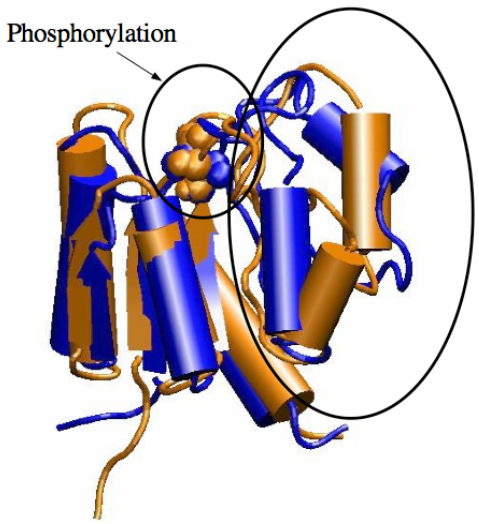


Figure 3: Graphic of protein being explored by the JHU group

During this first year the Open Science Grid has indeed established itself as a leading distributed infrastructure for the high energy physics community. OSG provided the worldwide LHC collaborations with more than 30% of their processing cycles worldwide. Their data distribution tests reached up to one hundred Terabytes per day transferred across more than seven storage sites. Physics results of the two experiments currently acquiring data at the energy frontier at Fermilab's Tevatron, CDF and D0, now routinely depend on the use of the OSG for parts of the data analysis and event simulation.

As an example of how the OSG organizational and computational structures can respond flexibly to changing needs, OSG members and staff contributed half the resources and throughput to DZero for the unanticipated reprocessing of a five hundred million event data set. This enabled new results in the measurement of the B sub-S osciallation frequency and the combined limit for Standard Model Higgs search, among others.

“On very short notice OSG provided resources not available within the Dzero Collaboration,” said Dmitri Denisov, co-spokesman of DZero. “Success in DZero’s presenting over 30 new physics results at the Lepton Photon 2007 symposium in Daegu, South Korea would not have been fully possible without OSG’s flexibility to apply substantial CPU resources to critical projects."

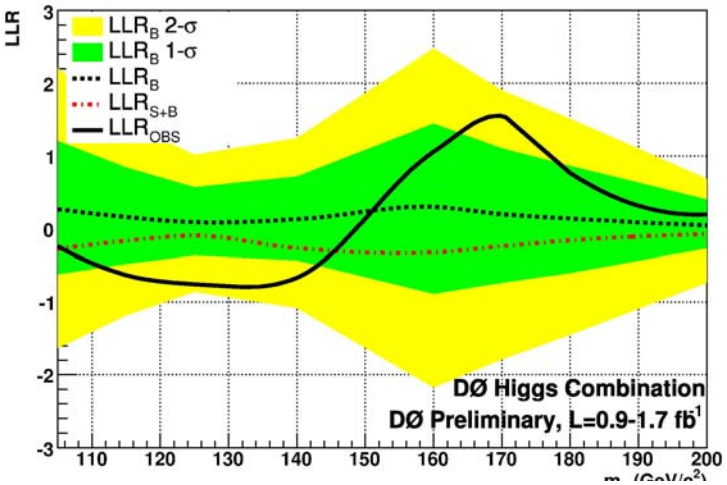


Figure 4: Log-Likelihood Ratio (LLR) distribution for the combined analyses over a possible mass range of the Higgs particle.

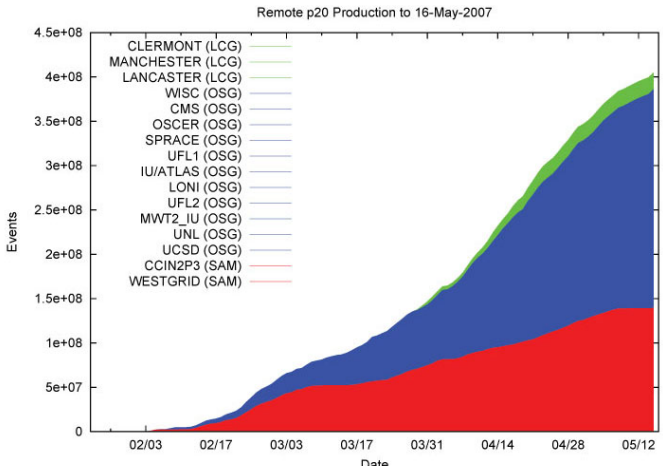


Figure 5: D0 event reprocessing production (OSG contributions in blue).

The Laser Interferometer Gravitational Wave Observatory (LIGO), one of OSG’s founding collaborations, is migrating its applications from the LIGO data grid to OSG. In an example of the common technologies and processes OSG facilitates, through collaboration with OSG and the LIGO Physics at the Information Frontier project, the DOE ESNET program has extended their Certificate Authority to support LIGO’s particular needs.

“OSG has essentially met the needs of LIGO for the ESNET Certificate Authority Upgrades” reports Patrick Brady, principal investigator for the LIGO PIF.

OSG is also supporting LIGO inspiral analysis application testing, which comprises workflows of more than ten thousand steps and thousands of jobs distributed across at least six OSG sites.

As part of broadening the nation’s campus integration into shared cyberinfrastructure, the OSG has helped provide access to and technologies for the first OSG-accessible Microsoft Windows-based farm at Clemson University and to several high-performance computing clusters at the NERSC shared facility at LBNL.

OSG has continued to focus on developing a robust, tested reference software stack packaged in the VDT and designed for ease of installation by users and administrators at new sites. As one of several areas of collaboration with TeraGrid, both projects have agreed to rely on the same underlying software versions of Condor and Globus, allowing interoperation.

“A key element of interoperation,” says Alain Roy, the OSG software coordinator, “was TeraGrid’s adoption of OSG’s build and test software”. As OSG adds patches, TeraGrid rebuilds Condor and Globus to stay in sync.

OSG has also continued its emphasis on enabling new communities to join and benefit from cyberinfrastructure. In collaboration with a number of universities, OSG has held grid schools at the University of Texas Brownsville (an MSI institution), the University of Nebraska Lincoln (in an Epscor state), and regionally at the University of Illinois,

Chicago. Our team of teachers has worked with roughly a dozen faculty members and ten times that many students to empower them in the basic use of the national grids.

Working towards global inclusion, OSG also contributed to the IceAge International School for Grid Computing and schools in Argentina and Brazil.

A typical student response came from Estelle M. Huff, a Ph. D. Candidate in chemistry at the University of Arkansas. “I don’t have much experience in programming If I can get a script together to submit jobs with chemistry software, I’m sure that other chemists would join the grid. I’m looking forward to being able to use the resources I’ve learned about here, both human and technological.”

Figure 6: Students at OSG’s Grid Summer School in Nebraska, August 2007

In summary, OSG has had a busy and productive twelve months, implementing a structured program of work across sixteen diverse institutions, engaging Consortium and external communities, extending international collaboration with Enabling Grids for E-Science (EGEE) in terms of shared software, security processes and architectural principles, and collectively responding to the dynamically changing priorities and needs of the growing global cyber infrastructure community.