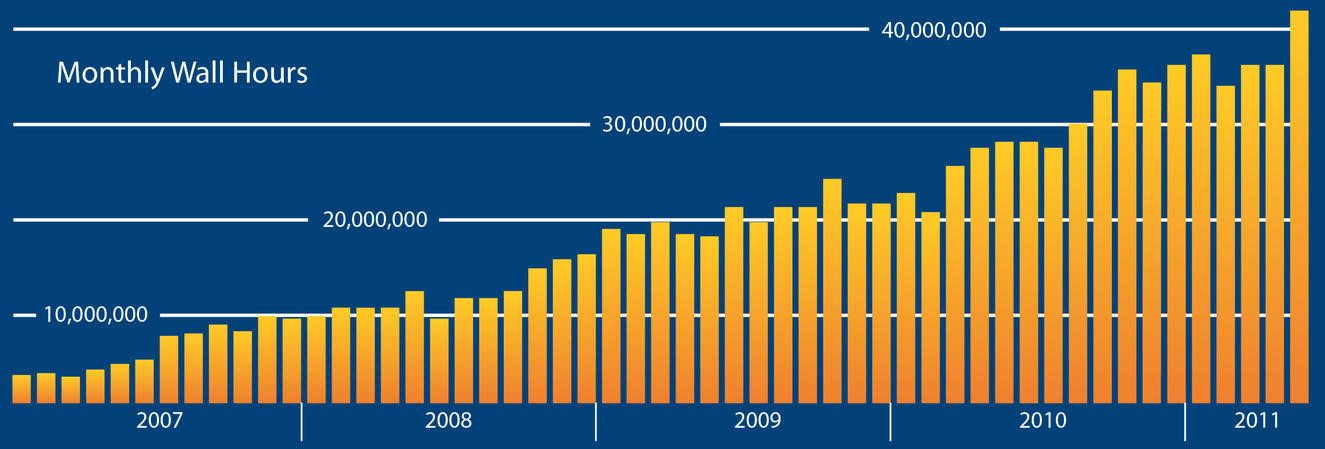
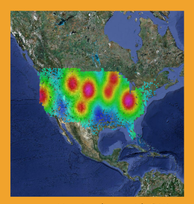
Open Science Grid

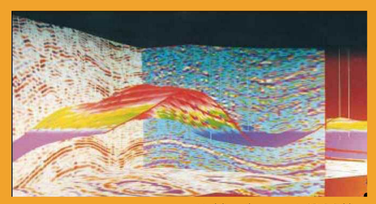
A shared national resource

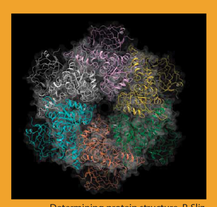
What is Open Science Grid?

The OSG is a U.S. multi-agency, multi-disciplinary national distributed computing grid designed to meet the needs of research and academic communities at all scales. In addition to providing governance and a common dependable infrastructure, the OSG distributes proven software solutions, offers support and consulting services, facilitates computer resource sharing, and organizes educational events.

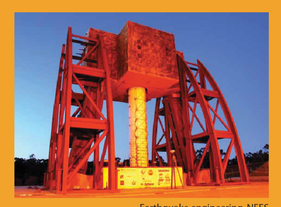



Mapping disease risk. S. Wang

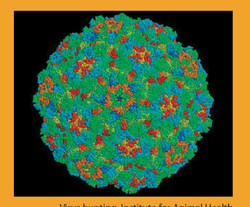

Modeling oil reservoirs in 3D. S. Siddiqui


Determining protein structure. P. Sliz


Enabling particle physics with the Large Hadron Collider. CERN

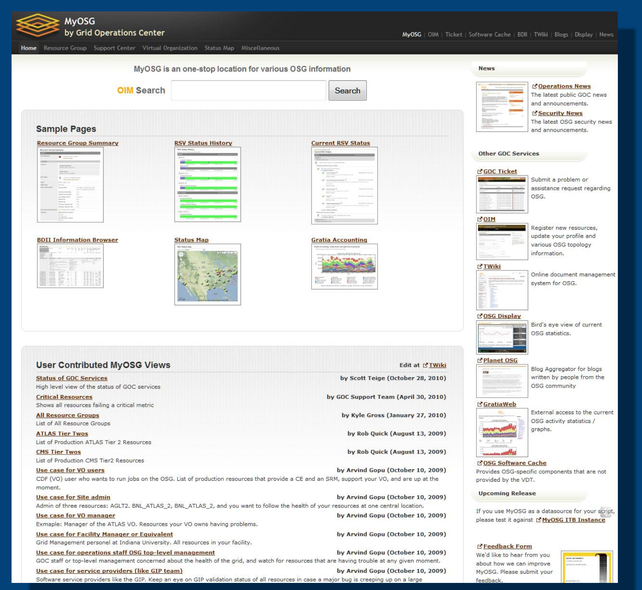

Earthquake engineering. NEES


Gravitational waves. LIGO


Virus hunting. Institute for Animal Health

Distributed Rendering Environment for educational games. IDEAlaboratory

MyOSG: A Portal for the OSG Community



MyOSG is a highly customizable web portal that allows vastly different categories of users to access information they find important to their role in a format that is convenient to them. The portal allows users to export and subscribe to a variety of information in a variety of formats, which they can view using tools such as iGoogle, iPhones, Androids, etc.

Fermilab: Proud to be a founding member of the Open Science Grid.
Contributing to DOE and NSF grid projects since 1999.

The OSG is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Partnerships for Science

OSG Partners and Projects

Advanced Networking Initiative
This project aims to prepare typical OSG data transfer applications for the emergence of 100 Gbps wide area networks.

Any Data, Any Time, Any Where
In partnership with other major Worldwide LHC Computing Grid stakeholders, we are developing the services to deliver worldwide dynamic data access.

CDIGS
Components of the Globus toolkit are included in the Virtual Data Toolkit. We collaborate with CDIGS to ensure communication of requests and issues so that as Globus evolves it will remain compatible with OSG software.

CorralWMS
We aim to integrate and enable the CorralWMS system across local and distributed computing resources as well as major national cyberinfrastructures.

CILogon
OSG is now testing site support for authentication provided by the CILogon service - an open source standards-based logon service that provides secure credentials for access to United States cyberinfrastructure.

ESnet
OSG depends critically on production networks operated by ESnet. OSG is working with the ESnet DOE Grids CA providers to transition responsibility for the capabilities, and we also work with ESnet on monitoring the end-to-end infrastructure.

European Grid Infrastructure
We continue to support interoperation with our peers in Europe. We work closely with EGI on operations, information, accounting, job and data exchange for VOs that operate across both infrastructures.

European Middleware Initiative
OSG continues to work with our peer software providers in Europe to support interoperation and use common software that meets the needs of both infrastructures.

ExTENCI
Through ExTENCI, we partner with XSEDE in order to develop and provide production quality enhancements to the National Cyberinfrastructure that will enable specific science applications to more easily access both OSG and XSEDE.

GridColombia
Our partners in Columbia are making good progress in establishing their own grid infrastructure. More than 10 sites have registered with OSG and attention is turning to helping new users.

GridUNESP
Our collaboration with the São Paulo State University's statewide, multi-campus computational grid in Brazil is reaching a level of maturity with several scientific application domains using the 6 installations locally.

High Throughput Parallel Computing
This project will enable our grid resources to run ensembles of small (8 to 64) way parallel jobs.

Internet2
OSG depends critically on production networks operated by Internet2. OSG works with ESnet on the deployment and use of Perfsonar to monitor data movement to and from the university sites.

Magellan
While Magellan was in operation US CMS and US ATLAS tested the cloud resources and LIGO and HCC VOs made good use of the computing cycles available.

Network for Earthquake Engineering Simulation
NEESComm provides the distributed IT infrastructure for the Network for Earthquake Engineering Simulation. Connections to OSG include Ruth Pordes as chair of the NEESComm cyberinfrastructure sub-committee of the project advisory committee.

OSG Summer School
An annual four-day summer school offering coursework on high throughput computing.

Pegasus
Pegasus is part of the Virtual Data Toolkit. LIGO, SCEC and other applications use Pegasus for workflow.

WLCG
You will have seen elsewhere in the Fermilab booth – clearly The Large Hadron Collider is a primary stakeholder of OSG via US ATLAS and US CMS and OSG contributes to the Worldwide LHC Computing Grid on their behalf.

XSEDE
OSG is a Service Provider to XSEDE, acting also as a partner in the NSF XD program. There are many touch points including operations, security, communities that use both OSG and other XSEDE resources. In particular, our Campus Grid activity works together with the XSEDE Campus Champions to provide researchers on campus access to local as well as remote resources appropriate to their applications and needs.

Table 2: Science Publications in 2010-2011 Resulting from OSG Usage

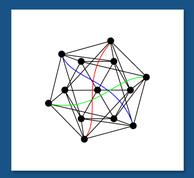
VO	# pubs	In Addition	Type of Science
ALICE	4		US LHC
ATLAS	38		US LHC
CDF	43		Tevatron Run II
CIGI	3		Geographic Information System research
CMS	57		US LHC
DES	1		Astrophysics
D0	23	8 accepted & 12 submitted	Tevatron Run II
Engage	20		Mathematics modeling virtualization research, molecular dynamics, protein folding.
GLOW	19		Neutrino physics, genomics, chemical modeling, molecular dynamics, proteomics, biology,
Grid UNESP	10	3 submitted	Cosmic ray physics, genomics
HCC	1		Protein modelling.
IceCube	7		Neutrino Physics
LIGO	11		Gravitational Wave Physics
Mini-Boone	1		Neutrino Physics
MINOS	7		Neutrino Physics
NYSGRID	1		Molecular dynamics
SBGRID	3		Structural Biology
STAR	9		Nuclear Physics
OSG & DHTC Research	2		Computer Science
Total	**260**		

Scientists and researchers from a wide variety of disciplines access Open Science Grid via virtual organizations to conduct their research. Last year, our virtual organizations reported a total of 260 peer-reviewed papers published.

www.opensciencegrid.org

University of Nebraska

The University of Nebraska and the OSG collaborate to expand available resources to Nebraska users and to further distributed computing research. OSG users opportunistically use Nebraska's computational and storage resources, and Nebraska users perform bioinformatics, medical, digital humanities, and computer science research on the OSG. Nebraska also investigates technologies and develops software for OSG. Here's three examples of research that Nebraska researchers are performing on the OSG.



Computational Proof of Graph Existence

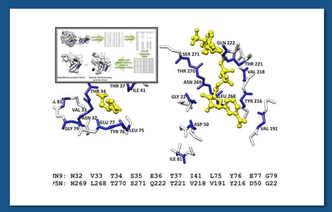
Derrick Stolee, a graduate student in mathematics at UNL is attempting to find or prove the non-existence of strongly regular graphs (SRGs) and directed strongly regular graphs (DSRGs) using discrete optimization techniques. The motivation for this exploration is that many important problems in modern mathematics involve the understanding of graphs, and that many tools and techniques have been developed to analyze them.

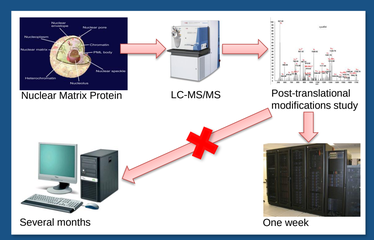
The current goal is to combine modern proof techniques with computational methods to solve unanswered questions related to complex graphs, including the existence of certain SRGs, space-bounded algorithms for reachability, and variants of the Reconstruction Conjecture. Stolee has already found a set of previously unknown DSRG parameters with no solution and another set of DSRG parameters with many solutions.

Comparison of Protein Active Site Structures

Robert Powers and his team at the UNL-based Power Research Group are leveraging OSG to aid in drug discovery using the Comparison of Protein Active Site Structures (CPASS) database and software.

The CPASS database is comprised of unique ligand-defined active sites – places on proteins where smaller molecules bind – that have been experimentally identified and stored in the protein data bank. The CPASS program compares these ligand-defined active sites to determine the similarity of their sequences and structures. CPASS will compare any set of ligand-defined protein active sites irrespective of the identity of the bound ligand. The Power Group uses CPASS in combination with their FAST-NMR assay to determine the function of an unknown or hypothetical protein, such as a potential new drug.





Mining Mass-Spectra Data for Medical Research

Protein post-translational modifications (PTMs) generate tremendous diversity, complexity, and heterogeneity of gene products, and their characterization remains one of the major challenges in proteomics. There are more than 300 different PTMs that are known to occur naturally in diverse biological systems.

Mass spectrometry is a central technology for identifying PTMs. Data analysis remains challenging because, when using conventional database-matching algorithms, the number of candidates to be tested expands exponentially as the number of modifications increases.

The Proteomics Collaboration at the University of Nebraska Medical Center, led by Shi-Jian Ding, has developed an iterative search algorithm, termed ISPTM, for identification of protein PTMs using the Open Mass Spectrometry Search Algorithm (OMSSA) database search engine. All OMSSA searches were performed using computing resources from the Open Science Grid and the University of Nebraska at Lincoln's Holland Computing Center. The preliminary data showed that the ISPTM algorithm was capable of identifying all the known PTMs at the proteome-wide level.

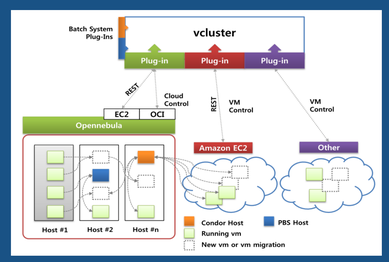
KISTI – 한국과학기술정보연구원 – Fermilab

In 2011, researchers at the Korea Institute of Science and Technology Information (KISTI) joined the Open Science Grid so that they could share computing resources with the CDF high energy physics experiment.

Today, KISTI has made hundreds of computing nodes available to the CDF experiment through OSG. Moving forward, the Grid and Cloud Computing Department in the Fermilab Computing Division and OSG User Support personnel have partnered with the Global Science experimental Data-hub Center (GSDC) team at KISTI to troubleshoot, test, and fine tune the OSG infrastructure for CDF production.

To get up and running on the OSG, representatives of KISTI visited Fermilab, where both the experiment and the Open Science Grid project office are located. During their visit, conversations between Fermilab and KISTI researchers inspired further collaborations.

Since then, KISTI software engineers have worked side-by-side with Fermilab personnel on FermiCloud, a project that deploys an Infrastructure-as-a Service platform to support the scientific computing program of Fermilab stakeholders. This has led to several improvements, including those profiled below.



vcluster: A System to Dynamically Extend Grid Clusters Through Virtual Worker Nodes

Grid resources have been expanded with worker nodes deployed dynamically as virtual machines at external clouds, such as Amazon EC2.

This diagram shows the architecture of vcluster, a virtual cluster system capable of utilizing computing resources from heterogeneous cloud systems. vcluster provides a uniform view for the jobs managed by the system. It distributes batch jobs to newly launched virtual machines over different types of cloud systems depending on the status of queue and system pool.

The main design philosophy behind vcluster is to be cloud and batch system agnostic; this is achieved by using a plug-in architecture.

Seo-Young Noh has been leading the vcluster effort.

InfiniBand Support on FermiCloud Virtual Machines

FermiCloud infrastructure has been enhanced with support for InfiniBand cards within virtual machines, enabling the deployment of high performance computing-like environments. This enables FermiCloud to support an environment to prototype MPI-based applications.

The project, in collaboration with KISTI, evaluated different techniques to expose the InfiniBand network cards of the hosting machines to the virtual machines deployed on FermiCloud. Currently, FermiCloud implements a software-based resource sharing technique (left diagram). Future work will deploy hardware-based resource sharing (right diagram).

Hyunwoo Kim has been leading the effort to provide InfiniBand support on virtual machines.