At-Large Virtual Organizations

Consortium Stakeholder Input To The OSG Council, and To The OSG Executive Board

Abhishek Singh Rana
Coordinator, Virtual Organizations Group

Ron Soltz & Latchezar Betev, LHC ALICE
Donatella Lucchesi & Rick Snider, CDF
Ion Moraru, CompBioGrid
Qizhong Li, Dzero
Nickolai Kouropatkine, DES
Zeno Dixon Greenwood & Horst Severini, DOSAR
Eileen Berman, Keith Chadwick, & Steven Timm, Fermilab-VO/FermiGrid
John Apostolakis, GEANT4
David Swanson, GPN
Russ Miller, GRASE
Jeffrey E DeReus, GROW
Stephen Moore & David A Cafaro, GUGrid
Steve Barnet, IceCube
Helio Takai & John Hover, MARIACHI
Gerhard Klimeck & Steve Clark, nanoHUB
Tom Furlani & Steve Gallo, NYSGrid
Piotr Sliz, Structural Biology Grid
Jerome Lauret, STAR

Open Science Grid

Official Report
Submitted: March 5, 2009

This page intentionally left blank.

Contents

DEADLINE: MARCH 3, 2009

Virtual Organization: Name

Stakeholder Scope: Science VO / Resource Provider VO or Composite VO

Mission Statement: Driving Force and Vision of VO

(Please feel free to include sufficient breadth and depth)

Activity by VO: Quantitative Metrics

(If VO is in start-up mode, please point it out, and provide "Future estimates")
(If VO is in full operation mode, you should list currently available numbers)

VO's Average OSG Utilization: Average CPU Hours/day, Average GigaBytes/day
VO's Peak OSG Utilization: Maximum CPU Hours/day, Maximum GigaBytes/day
VO's Resource Provisioning to OSG: Amount CPUs, Amount GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: (VO-specific metrics, and a detailed description)

VO Direction and Plans: Short-term 3-9 months

(If VO is in start-up mode, please indicate if targeted help and training - e.g., from Engagement, CampusGrids, Education groups in OSG - can be useful to accelerate)
(If VO is in full operation mode, indicate if activity will sustain, increase, or decrease)

(Estimate the "scale" of shift)

VO Direction and Plans: Long-term 1-4 years

(Please signify VO's long-term drive; Plans for growth; Value addition to OSG; etc.)

Needs and Expectations from OSG Consortium: Short-term 3-9 months

(E.g., Software; Operational; Security; Organizational; etc.)
(E.g., Guaranteed expectations - if any - of Resources, Throughput, Services)
(Make sure to mention timelines and priorities - if any)

Needs and Expectations from OSG Consortium: Long-term 1-4 years

(Please describe VO's projected long-term needs from OSG, in direct correlation with VO's long-term drive and plans for growth)

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(Mention any results that can help gauge OSG's impact on your *Science Portfolio* and productivity - i.e., from use of Resources, Services, Methods, Tools provided by OSG)

Miscellaneous

Virtual Organization: ALICE

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

To measure the properties of ultra-high temperature nuclear matter created in relativistic heavy ion collisions at the Large Hadron Collider.

Activity by VO: Quantitative Metrics

This VO is in start-up mode. The overall goal is provide full ALICE Computing production for the US institutions through full integration of ALICE grid tools (AliEn) with OSG. Current projected expressed in units of MSI2K are 1.6 MSI2K in 2009, growing to 9.2 MSI2K in 2013, transient storage needs of 0.8 PB (2009) to 3.8 PB (2013) and mass storage of 0.7 (2009) to 5.3 PB (2013). Bandwidth needs are approximately of 100/200 GB/day for average/peak utilization through 2012.

Activity by VO: Qualitative Science Value Output

Science Production: Peer reviewed publications and presentations at major scientific conferences with results from current ALICE data in proportion to the overall Ph.D. level of a high energy physics collaboration. Specifically, the VO should enable the US institutions to keep pace with current data reconstruction and in order to analyze data for potential new discoveries in the properties of nuclear matter at high temperature.

VO Direction and Plans: Short-term 3-9 months

VO is in start up mode to integrate OSG with existing ALICE grid tools (AliEn VO-Box). Immediate goals are to achieve full integration, large scale testing, and implementation at all US ALICE Computing sites.

VO Direction and Plans: Long-term 1-4 years

The VO should continue to keep pace with expected growth in ALICE computing resources.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

In addition to OSG job submission and monitoring, OSG will provide the reporting tools needed to meet Tier-1 or Tier-2 level of service requirements as specified in the Computing MOUs between CERN and each of the US computing sites.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

During a second phase, ALICE will proceed to test use of opportunistic resources across OSG sites. The level and conditions for opportunistic computing will be determined at a later date, by mutual agreement with ALICE, the computing sites, and their sponsors.

Virtual Organization: CDF

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

The goal of the VO is to allow members of the CDF Collaboration to perform all the computing activities relevant for the CDF Experiment and the analysis of data collected by it.

Activity by VO: Quantitative Metrics

Process hours from March 1, 2008 to March 1, 2009:
VO's Average OSG Utilization: 55.8k CPU Hours/day, 43k GigaBytes/day
VO's Peak OSG Utilization: 170k CPU Hours/day, 105k GigaBytes/day
VO's Resource Provisioning to OSG: 4440 CPUs, 0 GigaBytes

Note that the data consumption values, which include production dCache and data production volumes, are under-estimates, since some data consumption is from sources that are not tracked.

CDF supports no SRM storage, so has no storage available to OSG users not affiliated with CDF.

Activity by VO: Qualitative Science Value Output

Science Production: VO guaranties CDF users enough resources to analyze data collected by the Tevatron and to generate Monte Carlo events. Each year about 500 million simulated events are needed to perform data analysis.

The CDF collaboration publishes over 40 referred journal articles and typically ~250 conference presentations each year on the wide variety of physics where it tests the Standard Model from every available angle.

VO Direction and Plans: Short-term 3-9 months

See milestone (ii). Expect to start requesting and using opportunistic disk storage for MC staging pursuant to this goal.

VO Direction and Plans: Long-term 1-4 years

Demand for computing cycles at CDF will continue to increase slowly during the next few years of linear dataset growth, although some components of analysis CPU demand may decrease during this period as physicists leave the experiment. MC demand, currently the principal driver for off-site and opportunistic cycles, should remain approximately constant or experience small growth. The experiment will continue to invest in technology that will lower the barriers to efficient use of computing resources, both on and off-site. An effective strategy for moving MC data from off-site computing to on-site storage, for instance, may open the door to equally effective mechanisms for

remote data processing, a computing mode which is not currently part of the CDF computing model. Reduced barriers in this are may lead to some changes in the model.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Access to opportunistic CPU and storage. Continued support and enhancement of glide-in based overlay technology.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Access to opportunistic CPU and storage. Continued support and enhancement of glide-in based overlay technology.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(i) In 2008 CDF VO adopted the pilot-based Workload Management System (glideinWMS) as the submission method to OSG sites. Use of this new infrastructure system provides powerful scalability and makes Grid computing work as if in a local batch environment with the ability to handle more than 10000 running jobs at a time.

(ii) In 2009 CDF is planning to make opportunistic usage of disk space. Monte Carlo data produced using resources outside Fermilab are moved to tape storage in the Feynman Computing Center at Fermilab using old protocols that are not scalable. CDF wants to adopt a new method that consists of writing the output of jobs to the closest disk space using SRM tools available in OSG, and then organizing the movement to Fermilab in a way that depends upon the network and final storage space and bandwidth availability. This will allow the experiment to increase Monte Carlo production efficiency and CPU resource usage.

Virtual Organization: CompBioGrid

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

Our mission is to enable and support the use of OSG resources for computational cell biology applications – in particular modeling and simulation software of biochemical kinetics in spatially resolved geometries. The primary targets are users and applications associated with the National Resource for Cell Analysis and Modeling (NRCAM; www.nrcam.uchc.edu) and the National Technology Centers for Networks and Pathways (TCNP; www.ntcnp.org).

Activity by VO: Quantitative Metrics

The VO is currently in start-up mode. The numbers below are a "best guess" estimate of utilization within one year.

VO's Average OSG Utilization: 500 CPU Hours/day, 10 GigaBytes/day
VO's Peak OSG Utilization: 5000 CPU Hours/day, 100 GigaBytes/day
VO's Resource Provisioning to OSG: 440 CPUs, 5,000-10,000 GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: The first significant activity of our VO will be enabling of the Virtual Cell (VCell) application to use OSG resources. VCell (www.vcell.org) is a distributed model-building and simulation platform for cell biology that is currently served by a 4.6 TFlop / 28 TByte computational back-end at the University of Connecticut Health Center (the home of NRCAM, and of one of the five TCNP centers). As of February 2009, VCell had 2,224 worldwide users that have executed jobs, having 29,117 stored models and 160,539 stored simulations, of which there are 687 public (shared with the scientific community at large) models and 2,377 public simulations.

VO Direction and Plans: Short-term 3-9 months

We are currently struggling to get a small (60 CPUs) test CE resource site working properly. We plan to create and configure a SE site with support for scm after the CE site works properly. We will then grow the CE site with an additional 380 CPUs. Targeted help and training would be useful during this stage.

Once we have fully operational local CE and SE sites, we will be able to finalize development of the OSG interface for our applications. This is expected to take about 6 months. Only "manual-mode" job submission and usage of OSG resources will occur during that time period, with a low level of utilization.

VO Direction and Plans: Long-term 1-4 years

We expect to implement automated OSG resource utilization by VCell within one year. If succesfull, we expect to relieve current strict quota limitations for parameter scanning,

stochastic simulation runs, and implement large-scale parallel job capabilities, which could create a massive ramp-up of OSG utilization. Separately, we expect to implement OSG capabilities for new standalone applications under development at NRCAM, for example the Virtual Microscopy "suite", starting with the Virtual FRAP (VFRAP) tool which is currently in public beta stage. Lastly, we expect to support the other four TCNP centers to port and their applications to OSG and support their users.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

The only need in the short term is operational help with the design, configuration, and administration of our CE and SE sites (see section above on short-term plans). This is expected to be needed only during the next 1-3 months.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Satisfying the following long-term needs would have significant impact on our future utilization of OSG:

- support for long-running jobs (> 1 day, up to 2 weeks per job) – expected to be needed in calendar year 2011
- support for large scale datasets (> 0.5 TB at one site), both short-term storage (1-2 days) at job site and efficient transfer to our own SE site – expected to be needed in calendar year 2012
- support for small MPI reservations (tens of CPUs per job) – expected to be needed in calendar year 2012

We currently cannot make reasonable predictions of needs beyond calendar year 2012.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

2008 milestones achieved: pre-alpha porting of VCell jobs to be submitted to OSG CE sites (platform-specific compilation of customized stand-alone solver, customized input-output upload/retrieval – all dependent on manual submission)

2009 milestones planned: (i) deployment of own CE and SE sites; (ii) beta porting of VCell jobs (automated submission, user credential multiplexing); (iii) alpha porting of VFRAP jobs

Virtual Organization: DES

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

i. Dark Energy Survey Key Science

The observed cosmic acceleration or dark energy is a fundamental challenge to our understanding of gravity and the expanding universe. Many current projects seek to study the dark energy through its effect on the expansion history and growth of structures. Today, interesting constraints on the dark energy equation of state parameter w are only available when multiple, independent datasets are combined. Each dataset by itself provides only weak constraints on w, making it difficult to meaningfully test consistency of the datasets before they are combined. We have designed an experiment that will address the nature of the dark energy using four independent techniques. In the Dark Energy Survey, each of these techniques will deliver constraints on constant w models that are stronger than the best available combined constraints today. Combining these independent constraints will deliver w constraints at the percent level and provide a firm foundation for progress in our understanding of gravity.

The design driver for the DES is a multiband optical galaxy cluster survey coordinated with the South Pole Telescope mm-wave survey. The dataset required to deliver redshifts for the Sunyaev-Zel'dovich Effect detected clusters is also well suited for a weak lensing study and a galaxy angular power spectrum study. In addition, the non-photometric time can be used to carry out an extensive supernova distance survey.

The four key dark energy science projects are:

- A galaxy cluster study in collaboration with the South Pole Telescope cluster survey
- A weak lensing study of the fluctuation spectrum of dark matter
- A galaxy angular power spectrum study
- A SNe Ia study that will deliver ~2000 SNe Ia over the life of the project

Each of these projects will be the best in class on the timescale of this survey. In addition, each will deliver constraints on constant w models of 0.05 to 0.10. Together, these four experiments will deliver percent level constraints on a constant dark energy equation of state parameter w. Very strong constraints on the time evolution of the equation of state parameter are also possible. This will be the definitive dark energy experiment of its time, and it will lay the foundations for the much longer timescale Supernova Acceleration Probe (SNAP) and Large Synoptic Survey Telescope (LSST). A more detailed description of the Dark Energy Survey Science can be found in the document http://xxx.lanl.gov/abs/astro-ph/0510346.

This 5000 square degree deep g, r, i Z and Y band survey will be a rich dataset for many other science projects. The reduced individual images will be made available to the community one year after they are acquired on the telescope. Co-added images and object catalogs will be released twice: once halfway through the five year survey, and then again one year after the completion of the survey.

The Dark Energy Survey Data Management (DESDM) Project is developing the DESDM system, which will be used for the processing and calibration of the Dark Energy Survey (DES) data, and the DECam Community Pipeline, which will be used by NOAO to process DECam data obtained by non-DES observers. The DESDM team will also operate the DESDM system during the survey.

Activity by VO: Quantitative Metrics

VO's Peak OSG Utilization: Maximum 1485 CPU Hours/day, Maximum 0.25 TBytes/day

Activity by VO: Qualitative Science Value Output

Science Production: Our main activity is providing DES Data Management group with simulated images. For this we are running the simulation production once a year. With each year we increase complexity and accuracy of the simulated data as well as the volume of the simulated data. Our production cycles are followed with relatively long development cycles. With small test productions in between.

VO Direction and Plans: Short-term 3-9 months

Last year we have completed production of simulated data for DESDM data Challenge 4. This have produced 125 deg^2 of scientific images with total volume of the data 3.6 TB. The whole production cycle was done in 14 days using several OSG cites.
Right now we are in development cycle. We expect several small productions during the spring and summer development cycle, followed with main production in fall.
This year we plan to produce about 15TB of data using about 100K cpu-hours. We plan to use mostly Fermigrid. This will provide data for Data Challenge 5.

VO Direction and Plans: Long-term 1-4 years

In following years we will continue to provide data for Data Challenge 6 and 6.b increasing the data volume to 5000 deg^2, or 30 TB for DC6.b
Starting from 2011 till 2016 we plan to participate in seasonal data processing.
Each season data will be ruffly 50TB, and CPU needs are expected on the level of 10^5 cpu-hours.
We do not exclude other activity on OSG cites, e.g. specific physics interest groups can run their own productions not related to the main activity described above. We can not estimate needs of such groups right now.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

We plan to use increasingly the Fermilab resources. For this we are investing in storage and computing resources at Fermilab. Using other OSG cites probably will be limited to short jobs (< 4 cpu-hours) with limited storage needs. In our past productions the job preemption makes the process annoying, and we would like to avoid this.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

No estimations right now.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

In 2008 we have completed SDSS project. Two main data processing pipelines were running on OSG, and this contribute to the success of the project.
DES simulation was done for DC4 and we are preparing for DC5 in 2009.

Virtual Organization: DOSAR

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

DOSAR is a 'grass-roots' grid organization that focuses on community and campus based grids and promotes a wide range of interdisciplinary and educational activities within the organization and its member institutions. It is our mission to apply innovative grid techniques to High Energy Physics and to other disciplines in our communities.

Activity by VO: Quantitative Metrics

VO's Average OSG Utilization: Average CPU Hours/day, Average GigaBytes/day

2,000 CPU Hours/day, 100 GigaBytes/day

VO's Peak OSG Utilization: Maximum CPU Hours/day, Maximum GigaBytes/day

3,000 CPU Hours/day, 150 GigaBytes/day

VO's Resource Provisioning to OSG: Amount CPUs, Amount GigaBytes

750 CPUs, 5 TB

VO Direction and Plans: Short-term 3-9 months

(If VO is in start-up mode, please indicate if targeted help and training - e.g., from Engagement, CampusGrids, Education groups in OSG - can be useful to accelerate)
(If VO is in full operation mode, indicate if activity will sustain, increase, or decrease)
(Estimate the "scale" of shift)

- Deploy OSG and create Tier 3 sites at all DOSAR institutions
  - Active participate in computing grid efforts in LHC experiments
- Turn on new LONI_OSG site with storage and local DQ2
- Complete documentation for setting up D0, ATLAS and CMS computing sites
- Create Co-Linux OSG/CMS facility at University of South Alabama (and contribute to Co-Linux documentation)
- Document the Univ. Of Okla. Co-Linux setup.
  - Try out at UT-Arlington, other DOSAR institutions
- Develop innovative analysis techniques such as GPGPU and Virtualization
- Complete the deployment of SAMGrid at all DOSAR sites
- Continue supporting D0 MC and reprocessing efforts, leveraging all available computing resources
- testing of OSCER user MPI code on OSG? The user will work with the NERSC people to gridify their application. It would have to be someone

willing to try and break new ground. Potential benefits to the user are more resources to run on.

Much of our work would obviously be enhanced by close communication and involvement with the D0, ATLAS, CMS and Engagement VOs.

VO Direction and Plans: Long-term 1-4 years

(Please signify VO's long-term drive; Plans for growth; Value addition to OSG; etc.)

- Fully engage in LHC Collider experiments
- Continue participating in D0 MC and refixing according to our own priorities
- Play a leadership role in Grid computing in the corresponding states
- Extend our participation in other areas like education and outreach
- Collaborate in software development with other non-DOSAR institutions
- Leverage regional OSG support centers

Needs and Expectations from OSG Consortium: Short-term 3-9 months

(E.g., Software; Operational; Security; Organizational; etc.)
(E.g., Guaranteed expectations - if any - of Resources, Throughput, Services)
(Make sure to mention timelines and priorities - if any)

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Interaction with other VO's, particularly D0, ATLAS, CMS and Engagement
Possible support on Grid related proposals to funding agencies

(Please describe VO's projected long-term needs from OSG, in direct correlation with VO's long-term drive and plans for growth)

Significant Milestones (i) Met in 2008 (ii) Planned in 2009
Development of CoLinux Condor pool and incorporated in the OU OSG site
Plans for similar CoLinux farms at other DOSAR sites
OSG grid production at ISU and the University of Mississippi
Began support for new groups at the University of South Alabama and Susquehana University

(Mention any results that can help gauge OSG's impact on your *Science Portfolio* and productivity - i.e., from use of Resources, Services, Methods, Tools provided by OSG)

Virtual Organization: D0

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

The Dzero VO exists to provide computing for the D0 experiment, for data reconstruction, Monte Carlo generation, and physics analysis.

Activity by VO: Quantitative Metrics

(If VO is in full operation mode, you should list currently available numbers)

VO's Average OSG Utilization: 97160.2 CPU Hours/day, 227 GigaBytes/day
VO's Peak OSG Utilization: 190000 CPU Hours/day, 450 GigaBytes/day
VO's Resource Provisioning to OSG: 2300 CPUs, NA GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: D0 Data production, MC production and physics analysis.

VO Direction and Plans: Short-term 3-9 months

For data production, the usage will be increased by a factor of two in March-April, using D0 OSG resources.
For MC production, the usage will be similar to what we have been using.
For physics analysis, we expect the usage will increasing.

VO Direction and Plans: Long-term 1-4 years

For data production, the usage will be increased by 50% during next two years.
For MC production, the usage will be similar to what we have been using.
For physics analysis, we expect the usage will continue (slowly) increasing in the next 1-4 years.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

We would like to see better monitoring tools that provide status information about each phase of job execution from submission to completion. We also expect these monitoring tools to provide information about each site that includes total amount of available resources and the state of these resources, whether they are idle or occupied. These tools need to provide all this information in a reliable and timely (low latencies) manner and in a way that is accessible and easy to interpret.

We would also like to see a reliable resource selection service that automatically matches jobs to optimal sites with minimal input or intervention from the user.

We also need to know each site's policy towards each Virtual Organization (VO), e.g. whether all VO's have the same priority for their jobs, local storage, etc.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

DZero needs adequate access to the opportunistic resources available on the OSG for as long as the experiment lasts (2-3 years?).

- increased support for our VO at the various OSG sites (right now there are some sites that do not support our VO).
- hopefully the Consortium will continue to grow making more resources available to OSG users
- some way to link up with other grids allowing OSG users to be able to transparently tap into the additional resources available outside OSG such as LCG for example

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(i) in 2008, after D0 MC production coordinator informed the OSG team about the important role the local storage element plays in D0 MC jobs success rate, the OSG team immediately spent effort to installed the LSE at several sites. That has been a big help in increasing our MC production efficiency. (Thanks!) MC production has achieved record levels due in large part to production using OSG. DZero produced record levels of physics publications in 2008. This was enabled in part by the record MC production and the data production done using the OSG.

(ii) It is anticipated that DZero will again produce record levels of publications in 2009 using the MC and data produced using the OSG.

Virtual Organization: Fermilab VO

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

Our mission statement is the mission statement of Fermilab:
"Fermi National Accelerator Laboratory advances the understanding of the fundamental nature of matter and energy by providing leadership and resources for qualified researchers to conduct basic research at the frontiers of high energy physics and related disciplines. "

The Fermilab VO exists to provide all Fermilab employees and users an easy path to initial involvement with the grid, as well as to provide a unified VO structure for the smaller experiments and groups at Fermilab who do not have the manpower to do it on their own. It is also the sponsor VO of the resources that Fermilab provides to OSG via the FNAL_FERMIGRID and FNAL_GPGRID_1 resources.

The current subgroups of the Fermilab VO are the following:
Accelerator: Accelerator simulations for the proposed high-intensity proton Linac at Fermilab, code-named "Project X". Also cryogenic calculations.
Astro: Theoretical particle astrophysics. Currently inactive.
CDMS: Cryogenic Dark Matter Search, all data processing and simulation for this non-accelerator project is done on the grid.
Grid: Subgroup of Fermilab VO for general purpose use.
HyperCP: Fermilab Experiment E-871, search for CP violation in hyperons. Analysis almost complete. Reconstruction, Monte Carlo, and analysis were all run on the grid (and its precursors at Fermilab).
KTeV: Fermilab experiments E-799 and E-832, search for rare Kaon decay and measurement of direct CP violation in the K decay system. Analysis is almost complete, one analysis remains active. Grid activities include Monte Carlo simulation.
MiniBooNe: Booster Neutrino experiment E-898 at Fermilab, search for low-energy short-baseline neutrino oscillation. Grid activities are massive simulation to work out systematic errors.
MINOS: Fermilab E-875, long-baseline neutrino oscillation experiment with detector at Soudan mine in Minnesota. Reconstruction, Monte Carlo, and analysis all done on the Grid.
MINERvA: Fermilab E-938, near-detector neutrino-nucleus scattering experiment, now under construction. Simulation activities have begun.
MIPP: Fermilab E-907, Main Injector Particle Production—charged particle production experiment, analysis and simulation have been done on the grid since the beginning.
NovA: Fermilab E-929, Electron neutrino appearance experiment. Currently inactive.
NUMI: Simulations of improvements of neutrino beamlines and new neutrino beamlines for future experiments such as NovA and DUSEL.
PATRIOT: Used by the Fermilab-based maintainers of such Monte Carlo codes as Pythia
for validation purposes.
THEORY: Used by the Fermilab Theory group to run large-scale calculations of high-order loop diagrams, using tools such as MCFM and VINCIA.

TEST: Used by FermiGrid staff for internal testing purposes.

Activity by VO: Quantitative Metrics

Note, the storage element metrics are not yet available for the Fermilab SE, which is the only one that the Fermilab VO accesses with any regularity.
VO's Average OSG Utilization: 13,400 CPU Hours/day,
VO's Peak OSG Utilization: 115,000 CPU Hours/day
VO's Resource Provisioning to OSG: 1100 cpu cores, 6000 GB of dCache-based SE, 24000 GB of NFS-based $OSG_DATA area. FermiGrid Site Gateway also allows for opportunistic use of ~5500 job slots owned by CDF, ~5000 job slots owned by D0, and ~6000 job slots owned by CMS-Tier 1.

Activity by VO: Qualitative Science Value Output

Science Production: Science has been produced and published by a number of the Fermilab VO's user groups within the past year. We do not have a comprehensive list of papers available at this time, but know that there have been either publications or conference proceedings due to grid-related work from MINOS, MiniBooNe, CDMS, KTeV, HyperCP, and Accelerator. In addition, NUMI, Theory, PATRIOT, and MIPP groups are actively involved in work that will produce either publications or, in the case of PATRIOT, well-tuned Monte Carlo software that will then be used as a tool in LHC and ILC simulations.

VO Direction and Plans: Short-term 3-9 months

We are aware of the following plans of our user groups at the moment:

- Accelerator: Continued accelerator and cryogenics simulation at present level.
- CDMS: expect to continue analysis and simulation at present level.
- HyperCP, KteV: Both are essentially done, only clean-up processing.
- MiniBooNe: Are still taking beam and analyzing, expect usage at a similar level to last year.
- MINOS: Expect a large reprocessing of all their data in the coming months, a large increase in demand over 2008. Using GlideWMS they will be capable of running elsewhere than Fermilab, and high user demand may force them to do just that.
- MINERvA: Expect a MC effort to generate 10M events, also interested in getting some of their remote collaborating institutions to join the OSG.
- MIPP: No plan is in hand at the moment.
- NOvA: No plan is in hand at the moment, but expect one to be made at coming Neutrino workshop.
- NUMI: Recent influx of funding means a big increase in beamline simulation.
- PATRIOT: Continued tuning and validation of Pythia code at current usage level. Interested in running at other OSG sites.
- THEORY: Tuning efforts for Pythia-6.4 ("Perugia Tunes"). Development of the VINCIA antenna-shower and matching plug-in to Pythia-8. Color reconnection studies for top mass determinations at the Tevatron. Development of Pythia 6 and

Pythia 8.

In addition, the new MU2E, muon-to-electron conversion experiment at Fermilab will shortly become active as a new subgroup of the Fermilab VO. Other small neutrino experiments such as MicroBooNe will probably also have a need for access to grid resources. In addition, Fermilab is working on the Joint Dark Energy Mission which is currently creating an independent VO for the purpose of testing data handling pipelines. We continue to work with the Auger VO in hopes that they will join OSG as an independent VO.

VO Direction and Plans: Long-term 1-4 years

As Fermilab continues to push towards the intensity frontier, the experiments and accelerators currently under planning and construction will become reality. We expect the need for analysis and simulation to grow in a corresponding fashion. Given a recent ramp up to near full usage of FermiGrid itself in the last few months, we expect that there will be increased demand from on-site users to run off-site.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

The Fermilab VO continues to have a significant interest in the upcoming decommissioning of VORS and the commissioning of OIM and MYOSG. As a resource provider and service operator on behalf of the OSG we need a stable and well-documented API to the database. Also we expect that significant coordination with the GOC will be necessary as Fermilab transitions to the new version of Remedy Service Desk and the associated ITIL procedures and service catalogs that accompany it.

We are also dependent on many features of the Generic Information Provider software stack for our internal campus grid information system, but there has recently been a series of very productive discussions with the GIP developers to come up with a plan such that forwarding campus grid gateways such as ours can be handled within a branch of the GIP code itself.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

We expect that grid interoperability issues will be key on the 1-4 year time scale. There are a number of institutions in the above user groups who already have allocations either on European grids or on the TeraGrid who will want to do data processing on all of them. Therefore we expect to continue our interoperability efforts with those grids and with other campus grids in the OSG system, sharing what we have done. We expect to have an increased dependence on storage elements as well for data movement between Fermilab and our collaborating institutions.

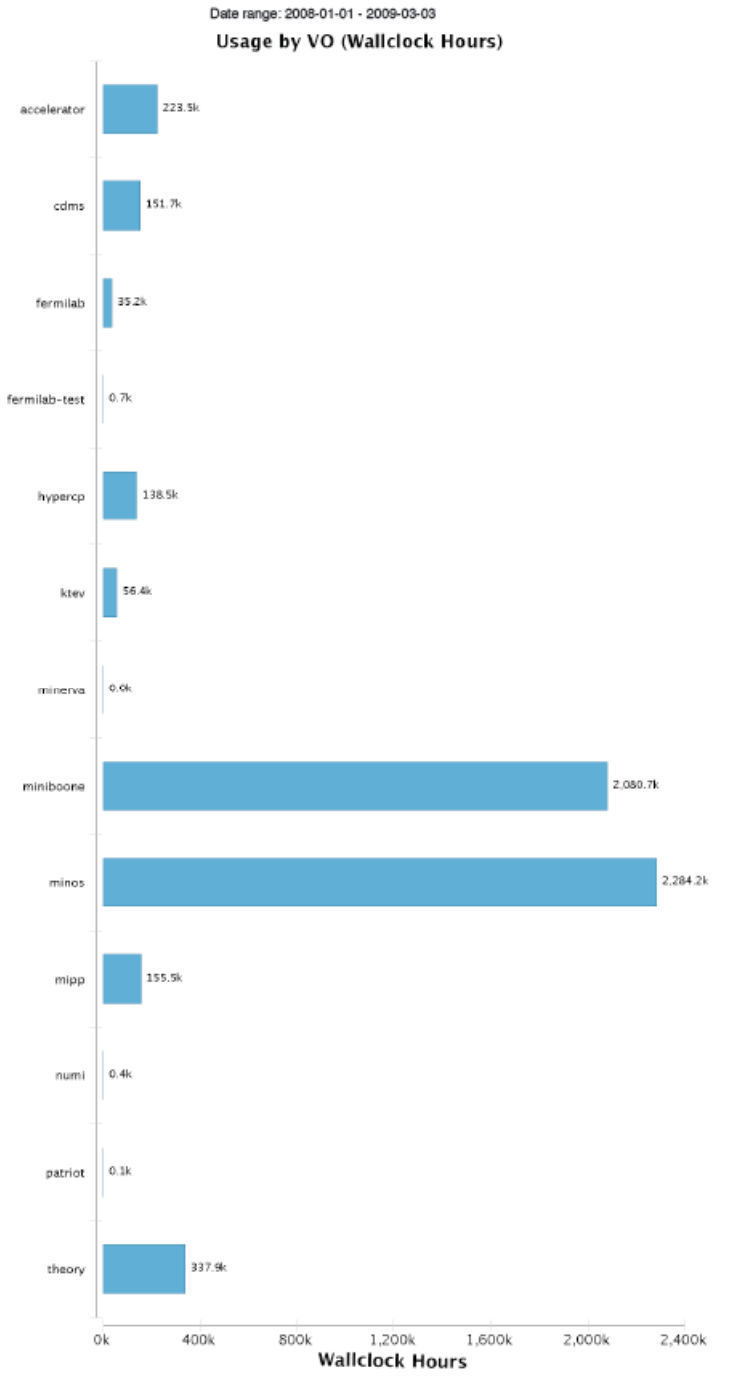
Significant Milestones (i) Met in 2008 (ii) Planned in 2009

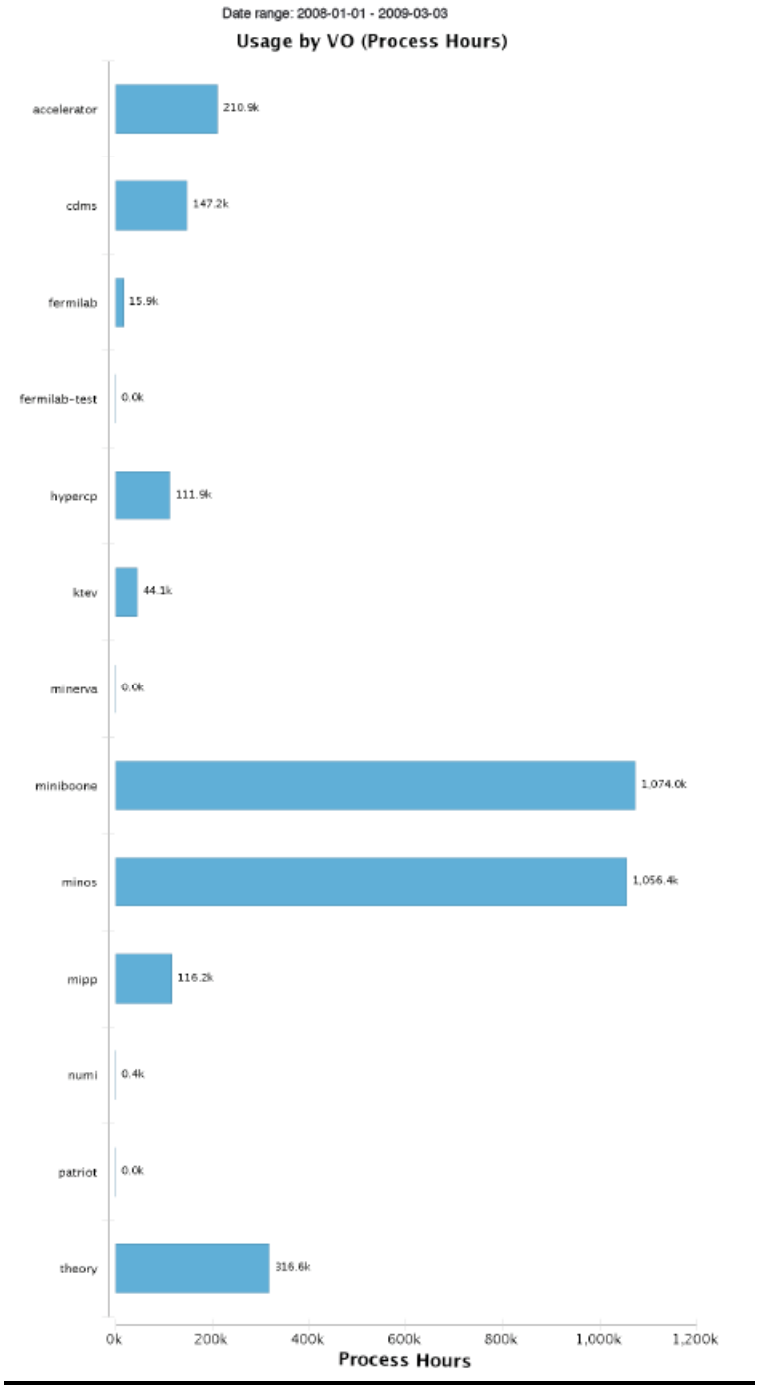
MiniBooNe, through opportunistic use of the FermiGrid, was able to accomplish in less than a week a cycle of simulation that had previously taken 2-3 months on dedicated resources. This simulation was crucial to announcing one of their results. There are

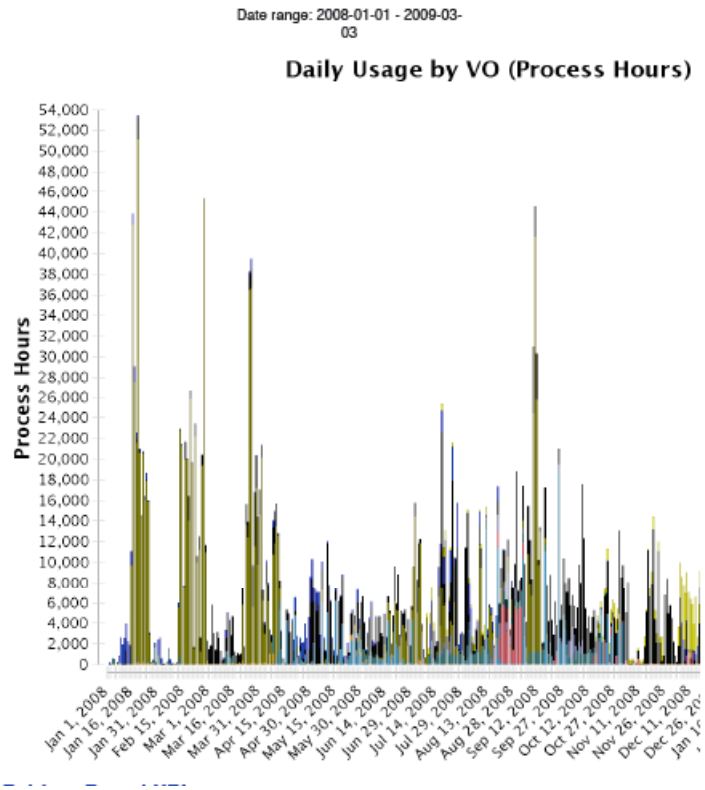
more milestones as well and we will be in better position to give these after the upcoming
Neutrino Computing workshop at Fermilab.

Miscellaneous

Graph is attached showing usage of Fermilab VO from Jan. 1, 2008 to present.




Date range: 2008-01-01 - 2009-03-03
Usage by VO (Process Hours)
accelerator
210.9k
cdms
147.2k
fermilab
15.9k
fermilab-test
0.0k
hypercp
111.9k
ktev
44.1k
minerva
0.0k
miniboone
1,074.0k
minos
1,056.4k
mipp
116.2k
numi
0.4k
patriot
0.0k
theory
316.6k
0k
200k
400k
600k
800k
1,000k
1,200k
Process Hours


Date range: 2008-01-01 - 2009-03-03
Daily Usage by VO (Process Hours)
Process Hours
54,000
52,000
50,000
48,000
46,000
44,000
42,000
40,000
38,000
36,000
34,000
32,000
30,000
28,000
26,000
24,000
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Jan 1, 2008
Jan 16, 2008
Jan 31, 2008
Feb 15, 2008
Mar 1, 2008
Mar 16, 2008
Mar 31, 2008
Apr 15, 2008
Apr 30, 2008
May 15, 2008
May 30, 2008
Jun 14, 2008
Jun 29, 2008
Jul 14, 2008
Jul 29, 2008
Aug 13, 2008
Aug 28, 2008
Sep 12, 2008
Sep 27, 2008
Oct 12, 2008
Oct 27, 2008
Nov 11, 2008
Nov 26, 2008
Dec 11, 2008
Dec 26, 2008

Virtual Organization: Geant4

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

The mission of the Geant4 VO is to validate Geant4 releases and Geant4 physics modeling. Currently the focus has been on the release validation. This ensures that the results of the physics modeling of Geant4 for High Energy Physics (HEP) experiment application only improves over time. It serves as a guarantee of HEP experiments can rely on the physics quality of Geant4 for their massive simulation productions. Even as refinements, fixes and new modeling approaches are implemented to improve certain aspects, the existing agreement with physics test beams already achieved is maintained and improved upon by avoiding significant side effects.

Activity by VO: Quantitative Metrics

VO's Average OSG Utilization: 120 CPU Hours/day, 1 GigaBytes/day
VO's Peak OSG Utilization: 5000-7500 CPU Hours/day, 40 GigaBytes/day
VO's Resource Provisioning to OSG: 0 CPUs, 0 GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production:

Geant4 is utilized in ATLAS, CMS, LHCb and other HEP experiments as the engine for detector simulation. The physics models utilized and their implementation is continuously developed and improved, addressing physics modeling deficiencies or gaps and fixing implementation issues or problems.
The regression testing undertaken by the Geant4 Virtual Organization ensures that existing physics performance for HEP calorimeter applications is not adversely affected by the revision of physics models, or changes in other parts of the Geant4 toolkit. It provides a significant degree of assurance that the physics performance, and reliability of Geant4, are maintained and improved, and that the large physics production of the LHC and other HEP experiments can rely on the physics of Geant4.

VO Direction and Plans: Short-term 3-9 months

The Geant4 VO has been in operation since 2005, providing regression testing validation Geant4 releases – with tests targeted for HEP applications. With the growth in the number of platforms and physics lists under test, the turnaround time for full testing has been growing during 2007 and 2008.

OSG resources and assistance will enable the timely delivery of the results of the large suite of regression test configurations, while addressing the further expansion of platforms to be tested. In 2009 these will include 32-bit and 64-bit Scientific Linux 4 (SL4= Red Hat Enterprise Linux 4), 64-bit Scientific Linux 5 (SL5) and 32-bit SL5 with gcc 3.4 and gcc 4.3.

Current plans are to utilize four platforms for each large set of 'runs', at least one each for 32 and 64 bit, for SL4 and SL5, for gcc 3.4 and gcc 4.3. This will increase approximately 50% the scale of testing, compared to runs of 2008. Plans to incorporate additional tests are under development, but are not expected to impact significantly on the computing resources required during 2009.

VO Direction and Plans: Long-term 1-4 years

In addition to the existing configurations used for regression testing, applications targeting the physics performance for different application domains are planned. The first extension planned will cover physics modeling for hadrontherapy applications. Use for large-scale productions for validation of new physics models are also expected.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

A set of small test productions during March and April 2009 are scheduled to test upcoming patch release and the monthly developments. Each 'run' will utilize a small set of test configurations (500 jobs x 4 configurations each). The results of the runs will serve to validate the patch release and to prepare for the large production runs of May. In addition these runs will be vital for preparing the ground for the large production.

Two major production runs are planned in 2009: May 20th – June 1st and November 23rd – December 14th. Each will include an initial phase of small sets of tests (each similar to a March/April run) for one baseline configuration and at least one release candidate.

Once a release candidate passes the small set, a large set of tests comprising of 1300 jobs per physics configuration (times 8-10 physics configurations times 4 platforms) will be run. Only these large sets will require the peak OSG utilization, typically over the period of 2-4 days for each production.

Additional runs, typically of small sets, are used to test late fixes, typically identified as a result of the large runs.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

During 2010 we expect a consolidation of platforms under testing to use only Scientific Linux 5, 32-bit and 64-bit, potentially with two different compilers (likely gcc 3.4 and 4.3). As a result the number of configurations to be tested is expected to be stable (at 4). Over the period 2010-2011 additional test cases (physics lists and new applications) are planned which will add to the number of configurations tested. Current predictions are of an increase of 25-50% in the number of configurations tested and CPU time required.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(Mention any results that can help gauge OSG's impact on your Science Portfolio *and productivity - i.e., from use of Resources, Services, Methods, Tools provided by OSG)*

During 2008 the regression testing for the June 9.2 beta and December 9.2 Geant4 release ensured the production quality of these releases, and identified five problems which occur with very low frequency.

During 2009 an expansion of the frequency and coverage of the regression testing is planned, the extension of platforms covered and the move to using the OSG. The addition of short intermediate productions between the scheduled releases of June and December, and the extension to additional platforms are the major developments.

Major milestones:

- Preparation of production capability on OSG, by May 6th.
- Regression testing of release candidates, starting May 20th and to be completed by June 1st, for the Geant4 beta release of June 5th.
- Regression testing of release candidates,, starting November 23rd and to be completed by December 14th, for the Geant4 release scheduled December 19th.

Miscellaneous

The assistance of the OSG deployment team, already underway, will be crucial for preparing and testing the machinery and undertaking the smaller productions in March/April in order to prepare for the first large scale production in May 2009.

Virtual Organization: GPN

Stakeholder Scope: Resource Provider VO

Mission Statement: Driving Force and Vision of VO

GPN is dedicated to supporting scientific research and education through the use of advanced networking technology and high performance computing. The vision of GPN is to enable the Great Plains region and its partners to lead in innovative learning/educational environments and collaborative research. The long-term goal of GPNgrid is to build a regional infrastructure for sharing computational resources across the region. To accomplish this vision, it is our aim to

- Partner to create a regional test bed focused on educational use;
- Interoperate clusters within and between campuses;
- Facilitate access to the grid for appopriate research problems.

Activity by VO: Quantitative Metrics

Start-up mode "Future estimates"

VO's Peak OSG Utilization: Maximum 2000 CPU Hours/day, Maximum 20000 GigaBytes/day
VO's Resource Provisioning to OSG: Amount CPUs, Amount GigaBytes – to be negotiated.

Activity by VO: Qualitative Science Value Output

Science Production: GPN will work with local researchers to port jobs, including parallel jobs, to the grid. Currently working with Engage team to help geoscience researchers run WRF. Idle resources will be available to OSG groups.

VO Direction and Plans: Short-term 3-9 months

VO is in start-up mode, please indicate if targeted help and training - e.g., from Engagement, CampusGrids, Education groups in OSG - can be useful to accelerate

An additional 5600 core cluster interconnected with IB is scheduled to be merged with current GPN resources at UNL later this spring. We will be looking for an initial surge of usage, as well other opportunities, at that time.

VO Direction and Plans: Long-term 1-4 years

VO's long-term drive; Plans for growth; Value addition to OSG
Longer term, GPN will be a mechanism to help grow a sustainable statewide grid for Nebraska, as well as to continue to pursue the potential of regional partners. With the addition of the new cluster, GPN will look to lower the barrier to entry to grid computing for a variety of application scientists, in particular those utilizing parallel computing.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

In the short term, the timing of the potential acquisition dominates GPN concerns. Once this occurs, it will be important to quickly facilitate a surge in usage, as well as to then facilitate increased grid usage by local users.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Maintainable storage solution. Initial assistance for novice users.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(ii) connection of multiple sites within the state of Nebraska; increase in resources by 5600 cores.

Virtual Organization: GRASE

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

The Cyberinfrastructure Laboratory (CI Lab) was founded at the beginning of the 21st century by Dr. Russ Miller, Senior Scientist at the Hauptman-Woodward Medical Research Institute and UB Distinguished Professor of Computer Science and Engineering at SUNY-Buffalo. Under Miller's direction, and with primary funding from the National Science Foundation (ITR, MRI, CRI, ANI), National Institutes of Health (Program Projects), federal and state appropriations, a Buffalo-based grid was developed (ACDC-Grid) as an experimental grid to serve as a platform for developing an institutionally-distributed platform integrating data and computation. The follow-on Western New York Grid (WNY Grid) was designed and deployed, with input and advice from members of the Open Science Grid, in order to aggregate expensive computational resources at institutions throughout Western New York. The overwhelming success of the WNY Grid led to the design and deployment of the New York State Grid (NYS Grid), a production-level grid based on the Open Science Grid software stack. NYS Grid was developed and deployed for multiple purposes, including advancing computationally- and informationally-demanding areas of STEM research, as well as education, outreach, and training of faculty, staff, and students at institutions throughout New York State. These institutions included K-16 academic institutions, non-profit institutions, research institutions, government agencies, and for-profit companies.

Members of Miller's CI Lab include(d) numerous high-school, undergraduate, graduate, and postdoctoral students, as well as SUNY-Buffalo technical staff, most notably, Jon Bednasz and Steve Gallo. Note that Mark Green was a key member of the CI Lab and was instrumental in leading much of the technical effort of the CI Lab during Dr. Green's tenure in the CI Lab, both as a post-doctoral associate and as a member of Center for Computational Research (CCR) during Prof. Miller's tenure as Founding Director of CCR.

During the development of the WNY Grid, the CI Lab saw an opportunity to educate non-physics scientists and engineers on the advantages of grid computing and cyberinfrastructure, as well as provide an opportunity for such research groups to utilize WNY Grid (later NYS Grid) and the Open Science Grid, which was significantly underutilized in preparation for the LHC coming on-line and welcomed significant utilization for the good of the scientific community and as a way to shake out OSG. Dr. Green worked with the Open Science Grid to establish the GRASE VO for the purpose of reaching out internationally to the non-physics community in terms of STEM-based research, education, outreach, and training.

The CI Lab worked with numerous community codes, assisting in their port to various grids, and performed significant outreach throughout New York State and Southern Ontario in terms of faculty, staff, and students – i.e., Education, Outreach, and Training in areas of grid computing and general cyberinfrastructure. This EOT took the form of working with system administrators in terms of bringing up clusters and connecting such clusters to the grid, to working with students in terms of education from the systems level through the algorithms level, to working with scientists and engineers in terms of a broad

understanding of grids and cyberinfrastructure, to working with graduate students, post-docs, and staff in research groups to port codes to our systems.

In addition, the CI Lab focused its internal research on cyberinfrastructure, in particular, grid-based areas that include, but are not limited to, Job Monitoring, Predictive Scheduling, Virtual Data Grids, an Operations Dashboard, the integration of Data and Computational Grids, Visualization, Application Templates, and Portals, to name a few. The current focus of the CI Lab is on mixed-mode computing for single-precision applications that take advantage of clusters of GPGPUs.

Activity by VO: Quantitative Metrics

Recently, the primary activities of GRASE have included Education, Outreach, and Training across New York State. This includes EOT efforts involving a grass-roots NYS initiative that appears to be waning after the steering committee made some unfortunate, unilateral, politically incorrect, and arguably unethical decisions on behalf of the group.

Activity by VO: Qualitative Science Value Output

Currently, the focus of the CI Lab's scientific effort is on applications that port to GPGPU-based clusters, including a 50 Tops cluster in its lab. The announcement of this machine is currently embargoed until a public announcement, which is currently scheduled for the middle of March, at which point the system will also be available to NYS Grid and Open Science Grid activities.

VO Direction and Plans: Short-term 3-9 months

The VO is focusing its efforts on mixed-mode high-end grid computing, including using grids that contain a heterogeneous mix of platforms, including GPGPU-based systems, which require distributed, shared, and fine-grained computing in order to optimize codes for such systems. The CI Lab continues its tradition of working with grids that contain a variety of clusters, data storage devices, visualization devices, sensors, and so on. The individual elements of the grids of interest to the CI Lab involve different chip sets, operating systems, vendors, and configurations.

VO Direction and Plans: Long-term 1-4 years

Due to the heterogeneous nature of the CI Lab's New York State Grid, which is 6-8 years old, and due to some unfortunate political maneuvering by some scientists and administrators at institutions throughout the state (c.f., http://www.nysgrid.org/main/about/steering.maml), the GRASE VO is now focusing on EOT and providing a heterogeneous set of resources for code development and optimization.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Currently, the CI Lab is comfortable in riding the wave of the OSG software stack. Note that OSG made decisions (likely valid) several years ago that forced the CI Lab to abandon some of its core research in predictive scheduling and monitoring as it was no

longer able to collect the data that was required to perform activities central to the lab and funding agents. However, efforts have been redirected into other areas of cyberinfrastructure.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

At the moment, GRASE and the CI Lab simply need to be kept in the loop so that we are aware of decisions being made or on the horizon at OSG and potential changes in directions of the software stack that might impact activities within GRASE and research being conducted in the CI Lab. Steve Gallo continues to serve as our conduit to OSG, having replaced Mark Green a number of years ago after Dr. Green took a unique opportunity that was presented to him in industry.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

In terms of a resource provider, in addition to facilities that continue to be available on the current New York State Grid (most notably a 2000 node cluster at the Center for Computational Research), courtesy of the CI Lab, GRASE will deploy a 50 Tops Dell/nVidia Tesla S1070 cluster that will be available to the OSG community by the middle of March, 2009. (This information needs to be kept under wraps until the formal announcement, even though the machine was delivered at the end of summer, 2008.) In terms of the CI Lab's GRASE VO, the focus will continue to be on the support EOT/development of our constellation of systems available throughout New York State. It is anticipated that Binghamton and Hauptman-Woodward will continue to be the major users of compute cycles through the CI Lab's New York State Grid (NYS Grid).

Virtual Organization: GROW

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

The GROW VO is in place in order to fill a need for a grid computing resource on the campus of the University of Iowa and to develop and strengthen Grid-based CMS collaboration with CMS Tier-1/Tier-2/Tier-3 centers. In addition, GROW will strive to provide a local grid environment in order to engage the local users on campus.

Activity by VO: Quantitative Metrics

VO is currently in start up mode. At this time it would be premature to attempt a projection of utilization numbers.

VO's Average OSG Utilization: Average CPU Hours/day, Average GigaBytes/day
VO's Peak OSG Utilization: Maximum CPU Hours/day, Maximum GigaBytes/day
VO's Resource Provisioning to OSG: Amount CPUs, Amount GigaBytes
Depending on local resource utilization, schedulers will allow for external user/group resource allocation.

Activity by VO: Qualitative Science Value Output

Science Production:

CMS Specific Applications:
Access to large data sets at a few central sites
Access to small data sets at many distributed sites
Ability to create jobs that run on these data sets
Ability to submit jobs and track their progress
Initially, GROW will be focused on CMS jobs.

VO Direction and Plans: Short-term 3-9 months

Targeted help needed for smaller VO osg grid utilization.
Targeted help needed for CampusGrids and Education.

VO Direction and Plans: Long-term 1-4 years

- Expand grid based cluster resources.
- Engage campus participants in expanding cluster with local external resources.
- Add sufficient capacity to enable different scientific groups access to GROW resources.
- Potential for engagement of resources for grid software development.
- Develop sufficient infrastructure to benefit Tier 2 community.
- Engage other institutions for collaboration opportunities.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Initialization of OSG for our VO.
Grant opportunity education for expansion of infrastructure.
Ongoing communication concerning resource utilization.
Collaboration methods with other educational institutions.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Continuation of OSG service.
Notification of possible software updates and modifications in the infrastructure.
Ongoing communication between both parties concerning resource utilization.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

- Local cluster setup initiated.
- Cluster setup completed. CMS computing environment and job scheduling is ready for production. Infrastructure proposed for OSG Consortium is prepared. The OSG utilization is to be initiated and tested for functionality, operation and ease of reach from/to external sources. Complete OSG functionality at University of Iowa is aimed to be achieved and maintained in 2009. The University of Iowa High Energy Physics (UIHEP) Group will be utilizing OSG for CMS analysis on the grid environment. This will improve the power and the speed of UIHEP for producing CMS analysis and anticipated early results.

Miscellaneous

Our development of the CMS Tier-3 center involves the following planning aspects:

1.) To keep track of cost/performance dynamics of computer hardware, and investigate and make decisions about hardware solutions;
2.) To maintain/upgrade the network connectivity between Grid sites and their network bandwidth to outside campus;
3.) To develop and strengthen Grid-based CMS collaboration with CMS Tier-1/Tier-2/Tier-3 centers;
4.) To seek resource sharing solutions (including external and internal grant applications) and investigate budget requirements of administering and operating a larger system.

Virtual Organization: GUGrid

Stakeholder Scope: Composite VO

Mission Statement: Driving Force and Vision of VO

To provide researchers at Georgetown University and partner organization access to high performance computation resources they may not normally have access to within the scope of their research grants and projects. GUGrid's primary focus is on BioInformatics and Cancer research, but other research area's within our community will be supported as well.

Activity by VO: Quantitative Metrics

The GUGrid VO is in "restart" mode. Could be considered in startup mode with a working VOMS server. Due to several factors we have been unable to take advantage of OSG in the past. We are putting renewed effort into becoming active again. Please see VO Direction and Plans below.

VO's Average OSG Utilization: TBD CPU Hours/day, TBD GigaBytes/day
VO's Peak OSG Utilization: TBD CPU Hours/day, TBD GigaBytes/day
VO's Resource Provisioning to OSG: TBD CPUs, TBD GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: In Startup mode so no qualitative science output

VO Direction and Plans: Short-term 3-9 months

In the short term we are working on setting up a computational resource on the OSG grid from our new shared cluster. We have also begun to plan for an on the grid teaching experiment between a students at a community college in California and researchers here at Georgetown. It maybe helpful to enlist some resources from Engagement and Education OSG groups to help with getting the remote campus students onto OSG. This would be a significant increase in our current OSG usage.

VO Direction and Plans: Long-term 1-4 years

As we are able to get more researchers comfortable using Grid resources for their science, we hope to make this a standard option for researchers in need of them. We primarily want to focus on driving BioMedical/Cancer related research, specifically researchers with new ideas who may not have the same level of financial resources as more establish researchers.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Primarily we need stable secure software to access the grid, and computational/storage resources which we may use from our VO.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Much of the same, as the short term with increasing ease of use in the software stack.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

None available.

Miscellaneous

In the past we have not done well in engaging our researchers and proving to them what benefits OSG engagement can provide. This is being addressed and we hope to become more active partners in the OSG consortium.

Virtual Organization: IceCube

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

IceCube is a neutrino observatory for astrophysics to be installed at the South Pole during Austral summers over approximately six years. Many parts of the universe are inaccessible for study using other types of cosmic rays: protons do not carry directional information because of their deflection by magnetic fields, neutrons decay before reaching the earth and high-energy photons are absorbed. IceCube will open unexplored wavelength bands for astronomy including the PeV (10^15 eV) energy region. IceCube will answer such fundamental questions as to the physical conditions in gamma ray bursts and whether the multi-TeV photons, originating in the Crab supernova remnant and near the super massive black holes of active galaxies (AGNs), are of hadronic or electromagnetic origin.

Activity by VO: Quantitative Metrics

The IceCube VO is in start up mode. Estimates are based upon current usage of local resources. Currently, IceCube does not provide resources to OSG, but intends to do so in the future.

VO's Average OSG Utilization: 1000 CPU Hours/day, 500 GigaBytes/day
VO's Peak OSG Utilization: 5000 CPU Hours/day, 2500 GigaBytes/day
VO's Resource Provisioning to OSG: 0 CPUs, 0 GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: The IceCube VO will begin its OSG work for production of simulation data. We hope to increase production by a factor of two or more. When this is working smoothly 50-60% of our analysis work, which generates the science results for many of IceCube's publications and PhD theses, will be moved to the OSG.

VO Direction and Plans: Short-term 3-9 months

The IceCube VO's short term plans are to make our data accessible to OSG sites and then to begin using the OSG for producing simulation data. The first step is basically complete, although we will need to add additional capacity to handle the additional workload. Within the next six months we would like to transtion from start-up mode into operations mode. We will likely seek help from the Engagement and Education groups to ensure the process goes smoothly.

VO Direction and Plans: Long-term 1-4 years

In the next two years, construction of the IceCube Neutrino Observatory will be complete. As we near completion of the detector, local resources are being priortized more heavily to the operational needs of making the raw data science-ready. We will

need to increase our usage of OSG resources to ensure that the science goals of the IceCube VO can be met.

Within the next year we plan to have our simulation production routinely making use of OSG resources. In addition, we may start moving some of our sophisticated users to the OSG within that year. In the following years, we intend to provide resources to the OSG community.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

IceCube will need basic organizational support and a small pool of resources (primarily CPU, but possibly opportunistic storage) to test our software and storage systems.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

In the long term, the IceCube VO will be able to make use of 3000-5000 CPU hours per day. It is unclear at this time whether we will require dedicated resources or if opportunistic usage will be sufficient. Given the relatively small needs we would project that there are sufficient resources available opportunistically. However, future science goals could change the workload significantly.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

- Basic engagement with OSG
  - Registration of IceCube VO within OSG
  - Engaged with OSG VO forum
  - Basic IceCube VOMS availability to OSG
- Expand OSG usage
  - Get basic simulation production running and tested
  - Scale up I/O capacity to handle additional workload
  - Work with sophisticated users to run science applications on OSG

Virtual Organization: MARIACHI

Stakeholder Scope: Science VO/Science Education VO

Mission Statement: Driving Force and Vision of VO

The recent observations of ultra-high energy cosmic rays(UHECR) (subatomic particles with energies equivalent to a 100mph fastball) present a mystery whose solution may provide insight into the origins and evolution of the universe. There are no known sources within our own galaxy or those close to us that could accelerate particles to these enormous energies, and yet interactions of such particles with the cosmic microwave background should prevent their propagation from greater distances.

MARIACHI (Mixed Apparatus for Radar Investigation of Cosmic-rays of High Ionization) will search for UHECR by detecting reflected broadcast TV or FM radio signals originating from distant transmitters. MARIACHI will confirm these signals using scintillator arrays to be built and operated by high school students and teachers. Radio and scintillator sites around Long Island will use a distributed computer system using grid middleware to collect, display, and analyze the data.

MARIACHI is a unique scientific project that integrates innovative research with education. MARIACHI's primary goals are to develop new detection and computing tools for studying UHECR while making these tools accessible to a diverse community of scientists, teachers, and students

Activity by VO: Quantitative Metrics

Mariachi is still working toward greater use of OSG software and the OSG Facility, but progress has been slow. Although we aren't using the facility, we are active users of the software infrastructure and the VDT. Our current status includes:

-- Our data acquisition system uses grid certificates, mod_gridsite, VOMS and GUMS to control access by remote data collection nodes.

-- We use a grid-cert authorized Wiki for collaboration.

-- An R-based cosmic-ray simulation application has been written, and the plan is to parallelize it for usage on clusters and grids.

-- We own two racks of 32 worker nodes each that should be moved into our machine room at Stony Brook University soon. These are in addition to our current 4 node quad-core cluster. When the additional systems are in place we will provide them for OSG usage via whatever version of the ITB is in use at the time.

VO's Resource Provisioning to OSG: 64 + 16 CPUs, 2.0 TB Storage

Activity by VO: Qualitative Science Value Output

Since we are not yet using the OSG facility, we have no specific OSG metrics.

VO Direction and Plans: Short-term 3-9 months

We recently had analysis results that we believe validate our detection method. This creates an urgent need to analyze a much larger body of data to confirm the initial result. This data analysis is of a scale that the use of grid-style parallel processing will finally be necessary for Mariachi.

Toward that end, we are currently writing the job-processing infrastructure that will let us do mass analysis--first on our small cluster (4 worker nodes), and later on OSG. As part of this process, we will necessarily have to fully deploy the globus gatekeeper at our site which will include registering it for others to use.

VO Direction and Plans: Long-term 1-4 years

Longer-term, provided the project acquires additional funding, Mariachi expects to expand both our owned resources and our use of distributed processing on OSG.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

We rely on OSG for software (globus gatekeeper, VOMS, etc.) and for monitoring service (VOMS).

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Our most significant need in the coming years will be an easily-used grid scheduling system to streamline online data analysis.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

(i) MARIACHI found possible evidence of the validity of our scientific goal using data collected with the help of grid middleware provided by OSG. (ii) In 2009, we expect to use distributed computing to analyze further data on a larger scale than previously.

Virtual Organization: nanoHUB

Stakeholder Scope: Science VO

Mission Statement: Driving Force and Vision of VO

nanoHUB.org was created by the -funded Network for Computational Nanotechnology (NCN). NCN is a network of universities with a vision to pioneer the development of nanotechnology from science to manufacturing through innovative theory, exploratory simulation, and novel cyberinfrastructure. NCN students, staff, and faculty are developing the nanoHUB science gateway while making use of it in their own research and education. Collaborators and partners across the world have joined NCN in this effort.

Activity by VO: Quantitative Metrics

nanoHUB usage of OSG has dropped off significantly in the past year. This was due to the low success rate achieved in jobs submitted. The nanoHUB applications that in the past submitted jobs to OSG have been using local Purdue resources in the interim while we work to improve the OSG success rate. In the first two months of 2009 nearly 9000 jobs have been run by these applications. These jobs have a low data requirement and a run time between a few minutes and a few hours.

Activity by VO: Qualitative Science Value Output

Science Production: (VO-specific metrics, and a detailed description)

VO Direction and Plans: Short-term 3-9 months

In recent months a OSG taskforce was formed to accelerate the process of improving the nanoHUB/OSG job submission success rate. To date a new testing and monitoring mechanism has been put in place to better understand the root cause of job submission failures. A web page has been put up to communicate issues between nanoHUB, sites, and OSG staff. The next phase is to test application codes on a subset of OSG sites. Particular attention will be paid to determining why an application fails (other than grid problems). Results of these tests will also be made available in web format. An attempt will also be made to utilize the local Purdue campus resources (very large condor pool).

VO Direction and Plans: Long-term 1-4 years

Increase both user and application base. This will increase demand for computational resources and provide an opportunity to increase OSG usage.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Continued existence of the nanoHUB task force – see through to completion the mission.

Virtual Organization: NYSGrid

Stakeholder Scope: Science and Resource Provider

Mission Statement: Driving Force and Vision of VO

New York State Grid (NYSGrid) is a collaboration among New York institutions to create an advanced technological infrastructure that will strengthen research and education capabilities across the State.

Activity by VO: Quantitative Metrics

NYSGrid can be considered in startup-mode as we are actively training and recruiting our membership.

VO's Average OSG Utilization: 8 CPU Hours/day, ? GigaBytes/day
VO's Peak OSG Utilization: ? CPU Hours/day, ? GigaBytes/day
VO's Resource Provisioning to OSG: 2100 CPUs, 2000 GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production: Initial users are running community and commercial software for Computational Fluid Dynamics, Geophysical Mass Flow, and Computational Chemistry research. We do not have detailed metrics on these users at this time. Our work has been in recruiting users to run on their applications over the grid and assisting them with porting their codes.

VO Direction and Plans: Short-term 3-9 months

NYSGrid is a regional grid that supports a plethora of scientific disciplines and users across New York State. Over the next 3-9 months, with the assistance of the OSG GOC, we plan on investigating and implementing the services necessary to function as a regional grid within OSG such as internal monitoring, job match-making, and resource monitoring. In addition to implementing these services, additional users and applications will recruited to make use of grid resources. These resources will initially be within New York State. Our goal is for NYSGrid to appear as a single resource to OSG both for resource users as well as the GOC.

VO Direction and Plans: Long-term 1-4 years

Help to define the role of a regional grid within the OSG. This includes resource monitoring, job submission to the regional grid, and policy development. We also plan to add users, institutions, and computational/storage resources as necessary to support the needs of our users. Campuses in New York have also begun to implement Condor pools and we would like to see these pools made available to researchers, although security and political concerns at local campuses prohibit this from happening today.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

Working with the GOC to identify and implement a subset of software necessary to manage a regional grid.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Working with OSG to develop policies as to how a regional grid interacts with the OSG.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

A number of users were brought onto the grid in 2008 (Andrew Schultz, Barbara Polkis, Ercan Dumplinar). These users were not able to perform their research without the availability of OSG and NYSGrid resources as their home institutions could not provide the necessary resources. In 2009 we plan on developing a training program for researchers within the state and to provide users with recommendations and support for resource identification, selection, data management, and job management within NYSGrid.

Miscellaneous

We also have plans to participate in efforts to improve the facilities for running parallel jobs (e.g., MPI) via the grid.

Virtual Organization: SBGrid VO

Stakeholder Scope:

a) **Structural Biology Virtual Organization (Science division) – nationwide community of ~2000 research groups that utilize macromolecular crystallography, NMR and electron microscopy in their research.**
b) **Boston Regional (Composite VO) – Boston community of research computing groups.**

The science and regional scopes are synergistic, but might require VO separation in the future.

Mission Statement: Driving Force and Vision of VO

- To develop computational portals that will allow structural biologists to submit refined computational workflows to the national grid resources. Portals will broaden access to grid resources and facilitate new scientific approaches.
- On the regional front SBGrid mission is regional a) awareness, b) speedy integration, c) interoperability and d) sustainability. There are currently several independent high-performance initiatives in Boston area. Our aim is to create an efficient workflow that will allow regional groups to federate with the NSF Cybeinfrastructure and to maintain the local grid infrastructure.

Activity by VO: Quantitative Metrics

(If VO is in start-up mode, please point it out, and provide "Future estimates")
(If VO is in full operation mode, you should list currently available numbers)

VO's Average OSG Utilization: Average CPU Hours/day, Average GigaBytes/day
VO's Peak OSG Utilization: 1958 CPU Hours/day, Maximum GigaBytes/day
VO's Resource Provisioning to OSG: Amount CPUs, Amount GigaBytes

Activity by VO: Qualitative Science Value Output

Science Production:
Macromolecular X-ray crystallography is a powerful method that is used to determine macromolecular structures at high resolution. The technique is employed in over 3000 structural biology laboratories, and in just 2008 over 2000 structures have been deposited to the structural biology-clearing house –the RCSB Protein Data Bank. Approximately 65% of all new structures were solved using partial information from models that were previously deposited in the protein data bank, but the remaining 35% of the structures were solved *de novo*, often using a more time consuming and laborious technique called experimental phasing – EP. We have identified some EP models that contain previously known structural elements, but were not utilized as molecular replacement models, often due to minimal or no sequence identity. We hypothesize that the Structural Classification of Proteins (SCOP) database, with over 100,000 macromolecular domains, could be used to determine structures of a subset of EP macromolecules. This would allow automation and acceleration of the structure determination process.

In order to analyze the general utility of our approach we have recently analyzed all 500 EP structures from 2008 and found that many of them include fragments that were previously available in the RCSB. Subsequently we have performed molecular replacement computations that included 10 models with high similarity to each EP structure along with 10 domains randomly selected from the Structural Classification of Proteins (SCOP) database. In approximately 30% of cases the SCOP fragment produced a molecular replacement hit clearly separated from the clustered decoy models. Figure one illustrates this for a newly discovered protein structure. We are now extending our analysis. We have selected 3000 unique and representative domains form the SCOP database and use all of them as a decoy set against the best models. This large-scale computation is currently underway and will be soon completed.

To date we have run this workflow over 600 times and utilized in excess of 20,000 CPU hours at four different Open Science Grid sites (Harvard, University of California at San Diego, University of Wisconsin Madison, and FermiLab). Some sample results and reports can be found at these URLs:

http://abitibi.sbgrid.org/~ijstokes/mr-jobs/2ets/2ets-summary.html
http://abitibi.sbgrid.org/~ijstokes/tmscan-jobs/mad/2-molrep-fast-10/plots.html

We have worked with a team from Open Science Grid to improve efficiency and reliability both of our underlying grid infrastructure and also this specific workflow. This has enabled us to consider more ambitious computational problems, and we are developing web-based interfaces to allow scientists associated with SBGrid to submit various computational jobs to Open Science Grid.

Qualitative output:

1) To analyze all macromolecular structures that have been deposited in the Protein Data Bank in 2008, and were determined by experimental phasing methods. To evaluate if the subset of those structure could have been solved by molecular replacement.

2) To implement a user portal that would allow structural biologists to perform single data file molecular replacement computations against a library of SCOP domains (fast mode with SCOP unique, and slow mode with SCOP full).

3) To develop a robust job submission mechanism that would utilize a local 114 node MacIntel cluster (to be integrated as OSG resource) and allow to submit to remote OSG resources when needed.

VO Direction and Plans: Short-term 3-9 months

Please see list of goals for 2009 (below).

VO Direction and Plans: Long-term 1-4 years

Integrated a majority of high performance computing centers in Boston area with OSG. Provide 5-10 refined OSG portals for structural biology computing.

Needs and Expectations from OSG Consortium: Short-term 3-9 months

- SBGrid-OSG Task Force, Phase II to be established in September
- Presentation by OSG members during OSG regional meeting in May, and quarterly meeting in March.
- Technical assistance as needed.
- Continued access to other VOs.

Needs and Expectations from OSG Consortium: Long-term 1-4 years

Same as above.

Significant Milestones (i) Met in 2008 (ii) Planned in 2009

- SBGrid Annual User Meeting attracted close to 300 participants.
- A regional VO seminar attracted IT professionals from Harvard, Northeastern University, Harvard affiliated hospitals and MIT. Miron Livny deliver a talk about OSG and advantages of federation with the national infrastructure.
- Successfully completed Phase I of the joint SBGrid-OSGWide Taskforce (please see the attached report).
- We have restructured SBGrid VO and are well positioned to move forward (Peter Doherty will be in charge of SBGrid VO maintenance and outreach efforts, Ian Stokes-Rees will continue his work on structural biology portals, and Ian Levesque will assist as needed). Part of SBGrid team that is not directly connected to VO efforts is currently being restructured, and will soon be able to provide more support to VO efforts.
- Majority of data analysis for the Molecular Replacement project has been completed.

Planned for 2009:

Task Force:
- Maintain high availability and reliability of SBgrid resources. Reopen Phase I if needed.
- Phase II of the SBGrid-OSGWide tasforce will be started in September. The goal is to increase utilization of the SBGrid resources (jobs from remote sites).

Educational:
- SBGrid Computing School and the regional OSG meeting is scheduled for May 2009 (school.sbgrid.org)
- The first quarterly OSG seminar is scheduled for March 11th (Miron Livny will talk about Condor). Second seminar will be delivered as part of SBGrid Computing School (Igor will talk about glide-in based WMS from CMS). Two more seminars will be scheduled for late summer and fall, and local members of OSG will present.

Hardware Integration:

- Peter Doherty will upgrade the core VO infrastructure and utilize it in deployment of additional regional resources.
- The 114 CPU MacIntel cluster will be integrated with the OSG and utilized as a primary resource for portal computations.
- 2-3 local computing groups from Boston area will integrate resources with OSG and configure cluster resources to interoperate within SBGrid VO (West Quad Computing Group is the first candidate).

SBGrid VO Infrastructure:

- We will deploy and maintain a system to monitor OSG resources in Boston area.
- We will establish a standard integration procedure and use it in integration of all regional resources.

Portal and Science Milestones:

- This is defined under a separate proposal. Briefly, we are working on three different molecular replacement project (blind search, focused libraries, parameter optimization). Once the MR project is completed we will switch focus to EM computations.
- Portal for MR computations will become operational by late July.

Miscellaneous

Please give us at least 30 days to complete reports of this nature.

SBGrid-OSGWide Taskforce Report:

(a) SBGrid Resource Infrastructure: The core site, SBGrid-Harvard-East, has been at almost 100% availability. All critical tasks have been converged to completion and to success, starting from a near 0% availability in September. Less than a handful of minor tasks remain, but these are optional and can be deferred. Recommendation is that the other cluster, SBGrid-Harvard-Exp, be limited to only job submission and Portal integration, and thus, not used as a core site. We owe success to excellent contributions by Terrence Martin (CE team lead) and Jeff Porter (SE team lead).

(b) SBGrid Science Application: Job submission and Portal work should be the focus for SBGrid in coming months. I am confident that SBGrid team now has a solid view of Portal integration plan and timeline. As part of Taskforce, after porting SBGrid's MolRep application from SGE to CondorG, and workflow integration with JobRouter, successful data pre-staging (GridFTP/OSG_DATA) and job submissions to local core site and to GLOW, UCSDT2, were conducted. Usage of FermiGrid was being delayed due to a bug in VOMS server version at SBGrid (details in spreadsheet) and SAZ access control. Bug resolution tracking has been in progress by the VDT team. We noted that SRM is not needed for SBGrid application in near-term. Greg Thain's excellent contribution was key to the work.

(c) SBGrid-wide Architecture and Modeling: At this stage, since we have converged on almost all "operational" tasks, I welcome an "architectural" discussion. It may begin open-ended (spreadsheet items: Resource/CE - 11,14 and ScienceApp - 6ii,6iii,6iv).

There is a need to draw out different models, analyzing costs/benefits of each approach. Primarily, with an aim to converge on concrete technological choices, to help implement Piotr's original vision. This will be important especially toward SBGrid's future expansion plans. E.g., Choice of JobRouter compared to GlideinWMS for (b); Extraction and deployment of a subset of OSG site services as core SBGrid central services, with need for only a limited subset of services to be deployed on new additional SBGrid sites for (a).

My sincere gratitude goes to all colleagues who were active in phase-1, to OSG Sites (Rob Gardner, Steve Timm); OSG Accounting (Phillipe Canal, Chris Green); OSG Storage (Alex Sim, Tanya Levshina, Neha Sharma); OSG Software (Alain Roy, Mine Altunay); OSG Information Services and Metrics (Burt Holzman, Brian Bockelman); VO Authorization Services project (Gabriele Garzoglio); OSG Engagement (Mats Rynge); VO Group (Britta Daudert); and to Rob Quick and OSG GOC for expediting and customizing the ticketing procedures.

With future upgrades to sites and further integration of the application, and with regular accounting of metrics-related data, it is likely that new bugs and issues may be uncovered. Resuming the use of normal GOC and RT ticketing procedures, and of public email-list discussions, is encouraged.

Together with Sebastien Goasguen (Coordinator, CampusCI) and John McGee (Coordinator, Engagement), it is my great pleasure to extend a formal welcome to the SBGrid team and project. We look forward to long-term strong partnership with SBGrid and to excellent contributions by Piotr Sliz, Ian Stokes-Rees, Ian Levesque, Alexan Mardigian.