Open Science Grid

OSG Year3 Planning Process & Schedule

Chander Sehgal
June 5, 2008

Process Overview

- Goals and Work Plans developed by Area Coordinators with support from Project Manager based on
  - Strategic Drivers from OSG-ET and their own experience and expertise with OSG and Grid Computing
  - Feedback from OSG DOE/NSF Review
  - Original OSG Proposed Work Plan
  - Process improvements based on year1 and year2 experience and feedback
  - Incomplete work items from current year
  - Schedule for year3 Planning
- OSG-ET provides strategic input, reviews high-level goals, and approves the WBS, budget, and staffing
- OSG Council endorses WBS, budget, and staffing plans

Planning Culture

- Provide feedback to each other and project management to improve the planning process; nothing is written in stone
- Develop work plans that “stretch” to achieve the important goals and not just what we know how to do
- Be willing to propose the termination of specific work if you do not believe it adds value to OSG; in your own area and in other’s areas
- OK to plan items that can’t get done in just year3; we want to capture the roadmap
- Document and “check-off” responsiveness to specific requests from Science Communities

List of OSG Area Coordinators

* 1.1 Software – Alain Roy (& Tanya Levshina)
* 1.2 Operations – Rob Quick
* 1.3 Integration and Site Coordination – Rob Gardner
* 1.4 Troubleshooting – Shaowen Wang
* 1.5 Engagement – John McGee
* 1.6 Campus Grids - Sebastien Goasguen
* 2.0 Security – Mine Altunay
* 3.0 Education & Training – Alina Bejan
* 4.1 Usability and Scalability – Frank Wuerthwein
* 4.2 Work Load Management – Maxim Potekhin (& Torre Wenaus)
* 4.3 Users Support – Abhishek Rana (& Chris Green)
* 4.4 Storage Management – Ted Hesselroth
* 5.1 External Communication & Coordination – Ruth Pordes
* 5.2 Metrics – Brian Bockelman (& Ruth Pordes)
* 5.3 Resource & Project Management – Chander Sehgal
* 5.4 Users & Community Communication – Anne Heavey
* 6.0 Science Goals and Requirements – Ruth Pordes

Year3 Planning High-level Schedule

Work Item	Completion Date
Document Key Project Drivers a. Goals & Plans of the experiments b. OSG Strategic Drivers c. Feedback from DOE/NSF review d. Input from Original Proposal of Work	June 15
Share Key Project Drivers with Area Coordinators to guide their goals; capture high level Area Coordinator Goals; complete review of these goals by OSG-ET	June 30
Year3 draft WBS, staff plan, budget plan	July 30
Year3 WBS, staff plan, budget plan approved by OSG Council	Aug 30
Draft SOWs ready for approval	Sept 30
Project Execution Plan Document Ready for OSG-ET Approval	Oct 15
Baseline Budget, WBS, Project Execution Plan	Oct 30
85% of SOWs ready for Institutional SRO sign-off	Nov 15

Approach to Work Items

- OSG Work items can be classified into four categories
  1. Maintaining and Sustaining the OSG
  2. Work directly supporting the Science Communities (e.g. LIGO, US-CMS, ATLAS, etc.) covering immediate and long-term needs
  3. Enhancing the OSG based on experience and feedback from stakeholders, technology evolution, and best practices
  4. Addressing the feedback from the Jan 23 NSF/DOE review

- Flag work items with these categories to enable traceability and provide handles for prioritization, if needed; evaluate total work program for coverage against these categories

Year3 Planning Mechanics

a. Capture the Key Project Drivers & Share with Area Coordinators as input and catalysts for their year3 goals
b. Area Coordinators work with Project Manager to document high-level goals for review and approval
c. Project Manager works with each Area Coordinator to detail WBS, staff, and budget across institutions
d. Project Manager prepares draft staff plan, budget plan, and WBS for review by OSG-ET and OSG Council
e. Project Manager write SOWs and baselines budget and WBS
f. Project Manager and Executive Director write project execution plan

Open Science Grid

Specific Inputs (a start)
for
OSG Year3 Work Planning

Concerns from OSG NSF/DOE Review

1. **OSG Requirements driven by LHC/LIGO requirements; demonstrate linkage:**
   1. Articulate dependencies between stakeholder and OSG deliverables and milestones
   2. Add columns to Work Items showing specific lineage of tasks to major stakeholders.
2. **Metrics that inform of the progress of OSG and contributions to science, especially LHC /ASCR etc.**
   1. Add metrics specifically for the LHC and LIGO
   2. Add metrics specifically for ASCR
3. **Scalability of operations model**
   1. Measure effort for various operations tasks.
4. **Ever increasing scope of software stack**
   1. Institute gate keeper process for including software and planning for releases.
5. **How to manage/prioritize ever increasing demands**
   1. Institute action item registers
   2. ensure good attendance to Executive Team meetings.
6. **Increase tracking of Impact on research**
   1. Attempt to better gather papers

(Analysis of Reviewer Comments by OSG Project Manager)

OSG year3 Strategic Drivers

- Define and monitor needs and schedule of WLCG, ATLAS, CMS, LIGO and other defined stakeholders.
  - Define and track explicit dependencies between OSG and Stakeholder plans.
  - A) Ensure the entirety of reporting to the WLCG is robust and accurate. Including site availability and reliability (of the infrastructure and VO based), accounting, operations
  - B) Reach usable opportunistic storage to enable US CMS, LIGO to use site not owner by the VO.
- Improve the ease of installation, configuration, and use of the OSG software.
- Improve the efficiency and usability of OSG infrastructure for existing and new members.
- Provide means to scale the infrastructure, as well as measure its scaling and performance envelopes.

Year3 Process Improvements

1. In year2 we started capturing External Project dependencies; continue, improve, and streamline

2. Manage Internal Projects more effectively; need to define/document our Software Process
   - Define Scope, Requirements, Design, & Milestones
   - Reduce re-work by doing effective reviews at the right time
   - Improved execution leading to delivery of output that matches requirements & scope

3. Integrate Education more tightly into the broader OSG

Milestones from OSG Proposal: Phase II: Months 19-36 (pg. 1 of 3)

Overarching goals are to complete the extensions needed by the initial stakeholders, increase the capabilities and usability for new communities and establish operating procedures for effective use of the infrastructure.

Science Goals:

- LIGO: Expand the user community and types of applications such that the entire LIGO/LSC (LIGO Scientific Collaboration) user base derives scientific benefit from transparent operations across LDG and OSG. Stochastic and some aspects of the burst data analysis pipelines will be extended run on OSG.
- LHC: Support for low and then high luminosity LHC Physics analysis. The LHC data volume explodes from less than 1PB to more than 20PB during this phase! User activity changes dramatically both in number and intensity of use. Reliability and robustness is likely to be the overriding concern for the LHC application communities.
- STAR: Support for user batch analysis on the distributed facility, object based and interactive analysis for the STAR collaboration. Scaling of the order of 10k jobs/day and beyond as well as a robust infrastructure will need to be reached by the second 1/3rd of Phase II.
- CDF: Analysis CAF infrastructure and data analysis applications on the OSG-CAF.
- D0: Initial support for user analysis on some OSG sites. Continued and expanded use of OSG for all stages of the overall analysis chain.
- SDSS: Continued use with scaled resource needs.

Milestones from OSG Proposal: Phase II: Months 19-36 (pg. 2 of 3)

Overarching goals are to complete the extensions needed by the initial stakeholders, increase the capabilities and usability for new communities and establish operating procedures for effective use of the infrastructure.

Facility:

- Reduce the "in-effectiveness" metrics of the Facility by 50%.
- Deploy distributed logging infrastructure and establish periodic analyses of cyber security audit logs.
- Support transparent data movement between OSG & EGEE.
- End to end monitoring and problem determination across the Grid infrastructures.
- Support transparent movement of data and applications across OSG & TeraGrid.

Education, Outreach and Training:

- Support of I2U2 VO(s).
- Add new modules as needed e.g. SRM.
- Create tutorial material to provide necessary prerequisites in LINUX and networking.
- Setup of student VO for independent research projects.
- Modularize courseware for self-paced delivery. Test and extend existing material.
- South African Grid site launched for scientific analysis.

Milestones from OSG Proposal: Phase II: Months 19-36 (pg. 3 of 3)

Overarching goals are to complete the extensions needed by the initial stakeholders, increase the capabilities and usability for new communities and establish operating procedures for effective use of the infrastructure.

Extensions:

- Expand the use of new capabilities introduced in the previous phase towards additional application communities.
- Expect to spend significant effort on understanding robustness, reliability, efficiency, and ease of operations issues for the new capabilities as deployed on the production grid as a result of work in the previous phase. Work with computer science partners on new releases that address these issues.
- Deliver a first production system that incorporates new network management tools into the distributed facility.
- Deploy a high-level user language to consistently express user interactive analysis and batch-based workflow.
- Expect to work on specifying requirements for new capabilities with new communities that started initial operations on the OSG during the previous phase.
- Deliver the final auditing system to be used by OSG for the remainder of this 5-year funding period.

US-ATLAS OSG Needs - 1

(from Michael Ernst)

- SRM V2 capable Storage Element
  - U.S. ATLAS is not mandating a single implementation/technology
    - SRM/dCache is used at some sites
    - Others are using BestMan/Xrootd
- Support for Physics Analysis (OSG Application Area)
  - PROOF
    - Multi-User support in general purpose Processing Farm environment
- Compatibility OSG/EGEE
  - Resolve conflicts/inconsistencies w/ middleware client tools between gLite and OSG
    - Have physicists setting up mixed client environments to work simultaneously w/ ATLAS tools (dq2 user tools), OSG services (gatekeepers, gridftp services, etc) and low-level data transfer tools (glite-copy, etc)
    - LCG-utils is now included in the VDT but it is not clear that for the tools that for the tools we use now (and in the future, e.g. catalog interactions w/ LFC) that dependencies are resolved/consistent.
      - Desirable to have an “lhc-client”

US-ATLAS OSG Needs - 2

(from Michael Ernst)

- Security
  - Many unresolved issues w/ Security
  - Integration of Security across various Services (Storage, File Transfer, Globus, Pilots)
- Ticketing
  - There are many Ticketing solutions
  - Missing uniformity (US ATLAS uses RT, WLCG has GGUS, OSG …)
- Service validation/checking
  - Like SAM but “smarter”

US-ATLAS OSG Needs - 3

(from Michael Ernst)

- Improved Displays of Gratia Accounting specific to the ATLAS VO
  - Provide U.S. ATLAS facility views which groups our Tier-1 and Tier-2 centers, and display the CE's associated with each group
  - Details to be provided bu U.S. ATLAS
- Metrics displays of RSV data showing time-dependence of site availability data
  - Nagios for example is lacking time-dependent graphs of status indicators
  - A current status of selected probes, for the set of CE's we closely monitor in the Facilities
  - Plots showing trends for specific probes over each of those time intervals (could be in RRD way, as done by Ganglia or Cacti)

LIGO OSG Needs - 1

(from LIGO/OSG Planning Meeting on 7 Apr 2008)

1. Warren will contact Mine and Doug Olson to continue the discussions of the LIGO security plan and AuthComm project. Requests for OSG work and participation will go through Mine, including any requests for GUMS extensions and support, and discussion at the Facility and Security meetings. OSG regards the LIGO implementation as an interesting and informative application of integrating new attribute technologies into the OSG infrastructure, and hopes to work closely with LIGO on leveraging effort and looking for reusable/common components.

2. Under Patricks guidance, Adam, Eva, Britta, will leverage the three organizations – LIGO, Pegasus development group, and OSG – for the project to produce science results from the Inspiral Analysis code running in a grid environment (expected to be 8-12 months). As more effort is needed or issues arise that impact to meet the goals in a timely manner all/any of these organizations can be approached for help.

3. We will follow up on OSG underreporting of LIGO usage accounting, especially at Milwaukee. (There have been email exchanges on this since our Monday meeting).

LIGO OSG Needs - 2

(from LIGO/OSG Planning Meeting on 7 Apr 2008)

4. LIGO needs for WS Gram deployment and support will be better understood, discussed and integrated into the OSG planning.

5. OSG will have an internal discussion of the technical requirements and impact of support for jobs running under BOINC as part of the strategic and work planning for year 3-5 of the project.

6. Data movement and placement for LIGO applications on the LDG and OSG infrastructures would benefit from further technical discussion between the groups. There are probably at least 3 separate discussion threads: LDR on non-LDG sites; scheduling of data placement and access for jobs running under BOINC; scheduling of data movement and access for the Inspiral Analysis application work.