

Open Science Grid Internal Review

Introduction

March 14, 2014

Lothar Bauerdick
OSG Executive Director



Purpose of the Review

Open Science Grid

- ◆ OSG will have to prepare a year-3-and-beyond plan, that will be reviewed by the agencies later this year

- ◆ This internal review will provide opportunity for the project to get feedback on status and input on strategies and approaches to the planning beyond year 2
 - ★ input to help us develop a convincing work plan and put us on a trajectory for sustaining OSG

- ◆ Consultant are asked to give us frank advice
 - ★ to the Executive Team and to the Area Coordinators
 - ★ from the perspective of the OSG consortium and stakeholders



Update on OSG in numbers

Open Science Grid

- ★ Institutes providing resources accessible through the OSG are members of the OSG consortium, governed by the OSG council
- ◆ ~ 690M hours of DHTC in last 12 months of which ~90M were provided as “opportunistic resources”
- ◆ The OSG Fabric
 - ★ ~120 sites
 - ★ ~100,000 cores accessible





OSG Project

Open Science Grid

◆ OSG Project was extended by 5 years (2012-2016) after the initial 6-year run

★ will have an agency review of the work plan for the 2nd half of the project

Program	Funds/Year
NSF OCI	\$1,000k
NSF MPS	\$2,750k
DOE OHEP	\$1,600k
DOE NP	\$50k
Total	\$5,400k

Area Manpower	FTE
Technology	8.6
Release Mgmt	1
Operations	9.2
Campus Grids	1.5
Networking	0.4
Security	2.1
User Support	2.45
Project Office & Communications	1.4
Total	26.65

OSG Staff	OSG Responsibility
Miron Livny (Wisconsin)	Technical Director
Lothar Bauerdick (Fermilab)	Executive Director
Chander Sehgal (Fermilab)	Project Manager and User Support Lead
Rob Quick (Indiana)	Operations Lead
Rob Gardner (U-Chicago)	Campus Grids lead
Brian Bockelman (Nebraska)	Technology Investigations Lead
Mine Altunay (Fermilab)	Security Officer
Ruth Pordes (Fermilab)	OSG Council Chair
Michael Ernst (BNL)	ATLAS planning interface
Frank Wuerthwein (UCSD)	CMS planning interface

Tim Theisen, Tim Cartwright (Wisconsin) Software and Releases



Main Areas OSG Delivers On

Open Science Grid

◆ Production/Operations

- ★ provide the OSG platform/eco system of services, sites, software for DHTC
 - ◆ infrastructure services, operations support, cyber security and incident response etc to enable VOs to run DHTC workflows and data systems across OSG sites (à la LHC)
- ★ a production quality HTC facility for a large & diverse community of researchers
 - ◆ built on harvesting resources opportunistically from OSG sites
 - ◆ delivering amongst others as a XD Service Provider, through XRAC allocations
- ★ provide other added values:
 - ◆ user and host certificates, software distribution services, network monitoring, ...

◆ User Support

- ★ consulting on technologies, architectures and user support
- ★ spreading knowledge on HTC as a science problem solver

◆ Technologies and Software

- ★ developing concepts and blueprints, deliver an evolving software stack
- ★ packaging, system testing, patching

◆ Campus Grids

- ★ OSG Connect service
- ★ Campus Infrastructure Community

OSG Strategy For Campus Grids

◆ Initial approach to Campuses

★ provide a software product (BOSCO, “myHTC”) to help users create larger, more inclusive Campus Grids

- ◆ goal: a downloadable self-installable and self-configuring robust tool to allow researchers HTC on whatever resources are easiest available to her/him
- ◆ BOSCO is a very successful technology, but mostly used by systems experts
 - ◆ most of functionality is “HTcondor with ssh” → HTcondor project
- ◆ have not managed to make it into a “myHTC” product yet
- ◆ Q if we should try again a thrust to really productize the “myHTC” product

◆ Recently developed another approach

★ support Campus Grids with the new OSG Connect service

- ◆ provide campuses with the services to connect to the OSG and a useful set of services that help setting up a campus grid or extending an application into a campus grid
- ◆ campuses don’t have to build their own — OSG is running those services for sites
- ★ initial set of services exists, two new campus grids being established
- ★ getting some traction also in Atlas and CMS as a solution for Tier-3 etc

OSG As a HTC Facility, based on otherwise idle resources

- ◆ Have built a facility (“OSG-VO”) and a set of services
 - ★ used by a quite large and diverse community of researchers and groups
 - ★ is being provided across the fabric, with resources harvested opportunistically
 - ★ with 90M CPU hours provided to users, this is largely successful!
 - ★ this facility is also the basis for OSG as an XD Service Provider
 - ◆ 30.7M hours to XSEDE users (of which however 28.8M went to one application)
- ◆ A number of questions that should find answers to
 - ★ how can we add a wider range of customers, applications and use cases
 - ◆ what additional services/capabilities are needed to make this an attractive facility
 - ◆ e.g. should we spend effort to provide application “front-ends” like science gateways?
 - ◆ e.g. what can/should we be doing regarding data access and data management?
 - ★ what capabilities are needed to make this “production quality”, accountable, predictable, effective
 - ★ started a discussion and some work how to provision resources to OSG users
 - ◆ add allocation-based resources (e.g. XSEDE or DOE LC facilities)
 - ◆ add resources provisioned from cloud providers
 - ◆ add resources that are available to OSG members on their campuses, connecting campus computing facilities into OSG



OSG Metrics:

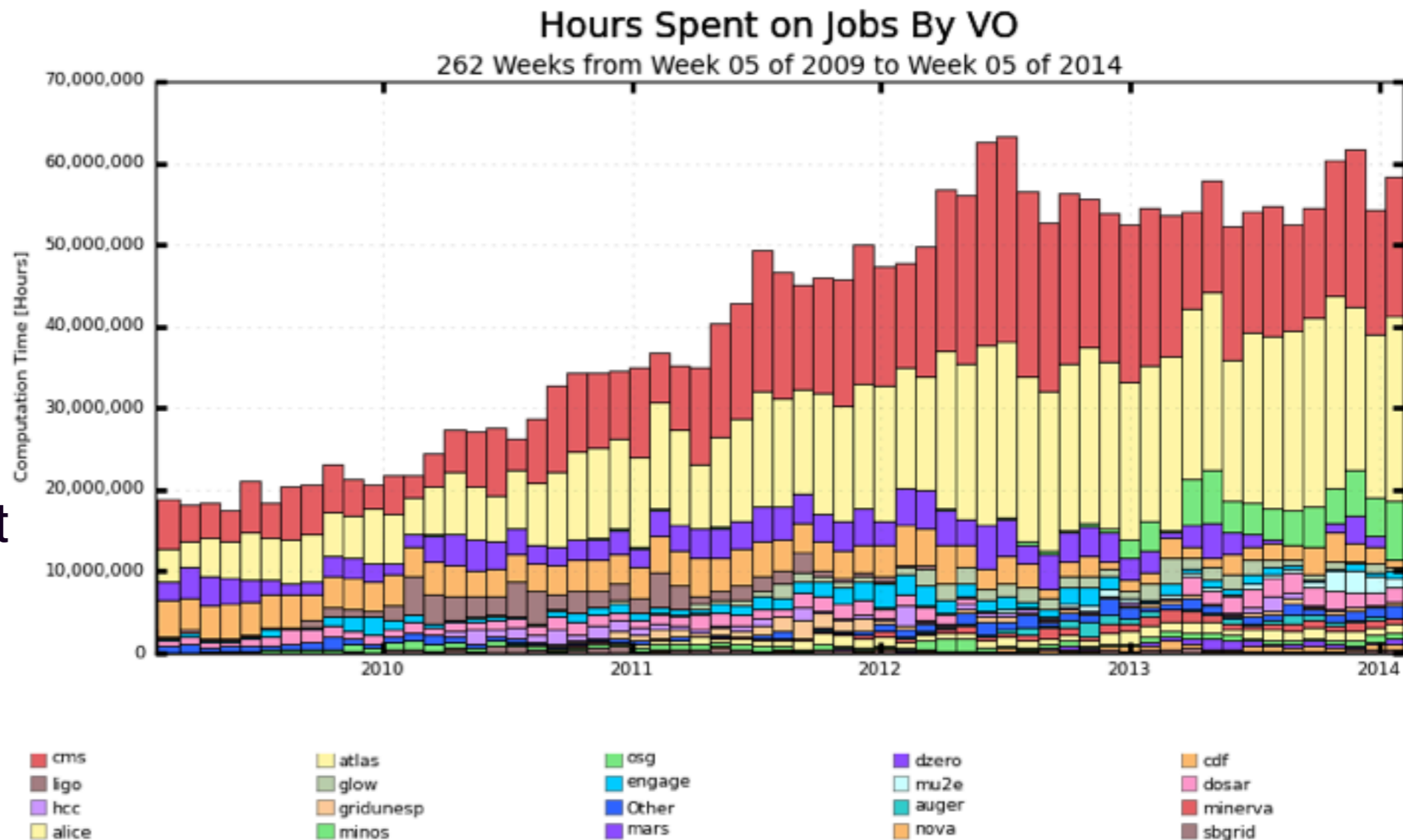
“Size” in Terms of CPU Resources

- ◆ We’ve leveled off, both in total resources used, and in non-LHC use
- ★ LHC will significantly upgrade its resources base in over the coming year
 - ◆ thus expect the overall use to increase again in 2015, a lot (70-100%)
 - ◆ this should also give an opportunity for additional “opportunistic” use

★ The Tevatron was “replaced” by other users, including other HEP and other opportunistic, but overall this component is ~flat

◆ should we worry about this?

★ can we do s/t about it?



Maximum: 63,246,967 Hours, Minimum: 2,829,821 Hours, Average: 40,397,651 Hours, Current: 58,263,696 Hours



Who's using OSG

◆ In CPU hours consumed, Atlas and CMS use 2/3rd of the resources

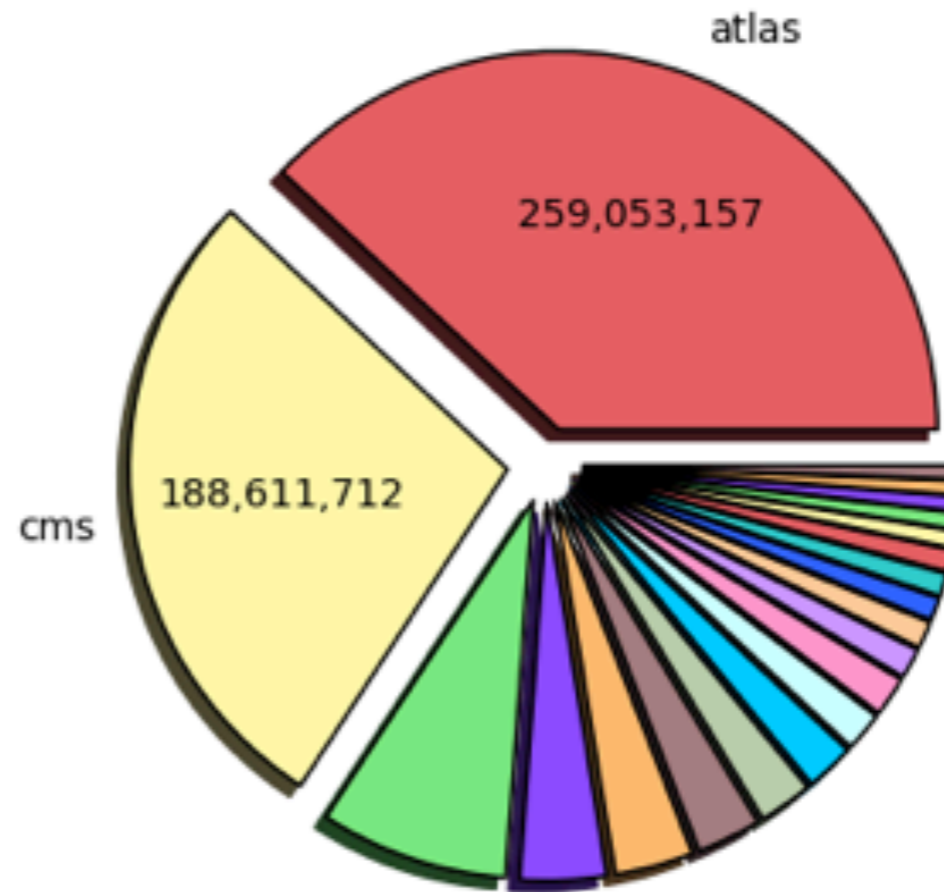
★ mostly on “owned” program-funded resources: LHC Tier-1 and Tier-2 computing centers

◆ third largest “user” is now the “OSG-VO”

★ comprising the bulk of opportunistic resources

★ 90M hours over the past year

Wall Hours by VO (Sum: 680,537,626 Hours)
52 Weeks from Week 05 of 2013 to Week 05 of 2014



atlas (259,053,157)	cms (188,611,713)	osg (54,981,660)	cdf (25,088,686)	dosar (22,942,960)
dzero (18,704,895)	glow (15,700,615)	alice (14,163,949)	mu2e (11,469,443)	minerva (11,457,000)
engage (8,725,196)	mars (7,881,645)	minos (6,679,094)	Other (6,631,880)	nova (6,368,366)
libne (4,919,198)	gridunesp (4,725,212)	auger (4,322,700)	sbgrid (4,103,054)	hcc (4,006,423)



OSG-VO providing resources harvested from sites

Open Science Grid

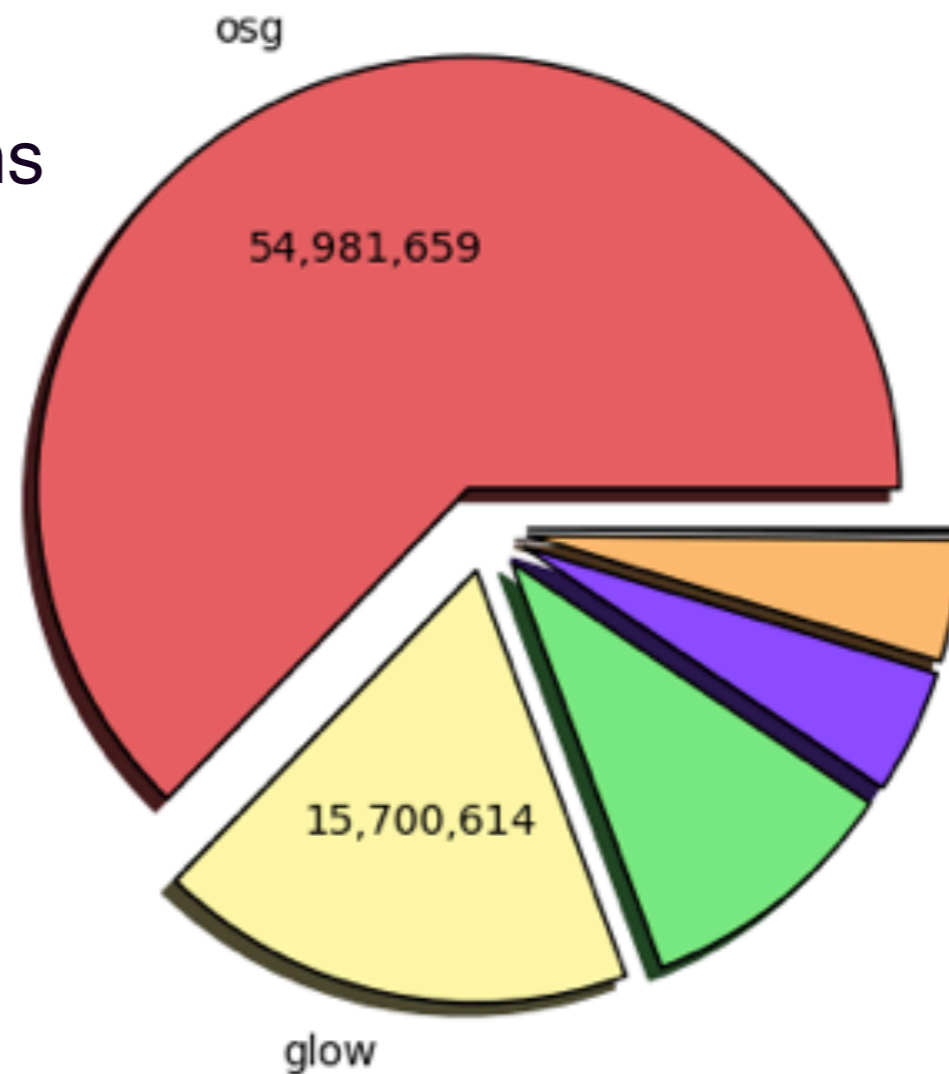
◆ OSG VO:

- ★ XD Service Provider
- ★ OSG Connect users
- ★ a number of science applications
- ★ individual science groups, who also come in through the engage VO

◆ Users flowing into OSG from campus grids

- ★ glow
- ★ hcc
- ★ sbgrid

Wall Hours by VO (Sum: 87,652,040 Hours)
52 Weeks from Week 05 of 2013 to Week 05 of 2014



osg (54,981,660)

glow (15,700,615)

engage (8,725,196)

sbgrid (4,103,054)

hcc (4,006,423)

gluex (135,092)



Opportunistic Resource Delivered To Individual Science Projects in 2013

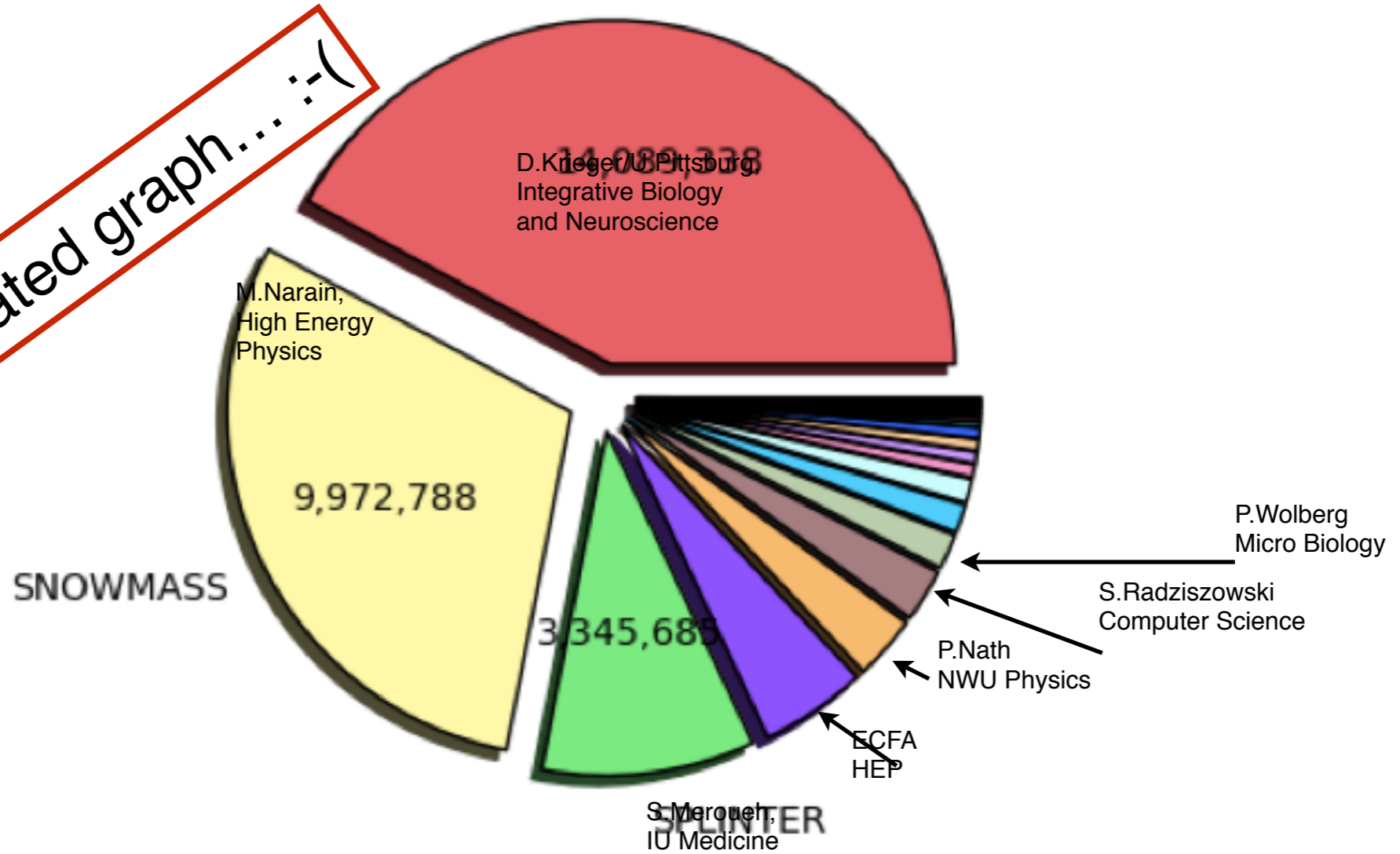
Open Science Grid

Wall Hours by VO (Sum: 33,432,599 Hours)

39 Weeks from Week 00 of 2013 to Week 39 of 2013

TG-IBN130001

Sorry for the outdated graph... :-)



- TG-IBN130001 (14,089,339)
- TG-PHY110015 (1,004,429)
- DUKE-QGP (371,289)
- Other (162,274)
- TG-PHY120014 (59,429)

- SNOWMASS (9,972,789)
- RIT (836,827)
- DETECTORDESIGN (202,613)
- TG-ATM130015 (77,169)
- CON-PSHMC (53,272)

- SPLINTER (3,345,686)
- UMICH (553,786)
- TG-DMR130036 (183,874)
- TG-DMS120024 (68,908)
- TG-ATM130009 (51,222)

- ECFA (1,669,414)
- TG-TRA100004 (444,375)
- EIC (175,818)
- TG-DMR120085 (60,625)
- OSG-STAFF (49,464)



Enabling Science — Metric

Open Science Grid

Collaboration/Campus/Community	Number of Publications
ATLAS	84
CMS	97
CDF	97
Dzero	26
NEES	1
STAR	15
GridUNESP	18
GLOW	61
UCSDGrid	4
HCC	1
University of Notre Dame	1
User Support	43
Total	448

- ◆ How do we improve in the metric of “number of researchers touched by OSG”?
- ◆ To achieve reach to science communities beyond the LHC, we believe the campuses are key
 - ★ e.g. through GLOW (U.Wisconsin), 48 UW researchers were enabled with access to DHTC/OSG
 - ★ e.g. HCC (U.Nebraska), ~26M hours to 23 OSG-XD users; and ~21M to 14 non-XD users
- ◆ Other approaches? E.g. Biology and Medical applications through Galaxy etc?
 - ★ should this be / can this become a focus of the project?



Other OSG “Value Propositions”:

Open Science Grid

◆ Software

- ★ organized bridging technology area and release team
- ★ how relevant is the OSG software stack, and how effective is the OSG-way of deploying it and making it available to the community?
- ★ beyond adding new software per stakeholder request, what do we need to do to keep the OSG software stack relevant?

◆ Networking Area

- ★ goal: push deployment of perfSonar instrumentation beyond LHC sites
- ★ OSG to provide extensive network quality information and monitoring capabilities

◆ Security

- ★ exploiting the existing trust relationships and traceability capabilities: how can we get the best value out of these capabilities for the community
- ★ should we put more or less energy into PKI, identity provisioning
- ★ more collaboration with XSEDE, Globus, etc on IDM?



Goals/Charge for this Review Meeting

Open Science Grid

- ◆ we invite you to interact with presenters, to address the charge
 - ★ are we meeting the needs and expectations of our stakeholders
 - ★ what is the quality and cost effectiveness of the services
 - ◆ how can we maintain or improve quality and efficiency
 - ★ what are the areas in which OSG provides most value
 - ◆ give advice how we should further develop these
 - ◆ give advice about extending the user base and reach
 - ★ advice towards extending the provisioning of opportunistic computing cycles
 - ◆ for individual scientists and research groups
 - ◆ for scientific applications
 - ◆ for connecting campus computing to OSG
 - ★ what new approaches and strategies should OSG develop, how should we prioritize
 - ◆ are there areas in which we should get more involved, like e.g. data/storage
 - ◆ should we prioritize areas above others or refocus our efforts
 - ★ ideas how to promote the adoption of high throughput computing in science



Agenda/Time Table

Open Science Grid

- ★ We start with a set of presentations from all OSG areas, ~20mins each
 - ◆ User Support: Chander Sehgal
 - ◆ Campus Grids: Rob Gardner
 - ◆ Operations: Rob Quick
 - ◆ Technology: Brian Bockelman
 - ◆ Software, Release Management - Tim Theisen, Tim Cartwright
 - ◆ Security - Mine Altunay
 - ◆ Network Monitoring Shawn McKee
 - ★ around noon: Open Q&A Forum
 - ★ lunch and caucus of reviewers
 - ★ ~ 2pm close-out from Reviewers
 - ★ 3pm Adjourn
-
- ◆ Thank you for taking the time and effort to help us with this!