

Open Science Grid

Workplan for 2nd half of OSG project and Agency Review

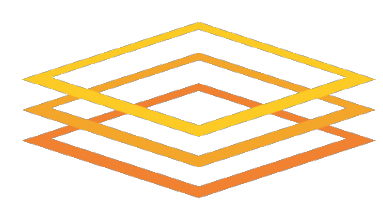
OSG Council Meeting
Norman, OK — October 22, 2014

LATBauerdick
OSG Executive Director

Agency Review

August 19-20, 2014

- ◆ Review of the workplan for the 2nd half of the project
 - ★ a series of preparatory internal reviews, meetings, discussions
- ◆ Prepared and submitted a number of documents ([link](#))
 - ★ detailed reports and a summary of year 1/2
 - ★ detailed work plan for year 3-5
 - ★ personnel, including CVs and role in project
- ◆ Charge was available only verbally
 - ★ a written charge (only NSF!) was eventually given during the meeting
 - ★ we were then asked to address all the charge elements in the Q&A
- ◆ Reviewers with backgrounds in HEP, NP, CS
 - ★ Tom LeCompte (ANL), Barbara Jacak (SUNY, Stony Brook), Craig Tull (LBNL), Tomasz Skwarnicki (Syracuse U.), Gordon Watts (Univ, of Washington), James Myers (U. of Mich.), Nicole Radziwill (James Madison U.), Paul Sheldon (Vanderbilt)
- ◆ Participation from DOE and NSF program managers
 - ★ including HEP/particle physics, ACI, nuclear physics
- ◆ Special 1h discussion on “OSG beyond year 5”



Agency Review, continued

- ◆ The review was of the workplan for the 2nd half of the OSG project
 - ★ meet the growing and changing computing needs of U.S. LHC program
 - ★ facilitate the increasing demand for High Throughput Computing (HTC) by the U.S. research communities at DOE labs and across university campuses
- ◆ OSG offers a diverse fabric of services that advances scientific discovery through improvements in computational throughput
 - ★ **Distributed HTC**, the shared utilization of large ensembles of autonomous resources toward a common goal, where all elements are optimized for maximizing throughput
 - ★ OSG provides a powerful **platform and ecosystem of DHTC capabilities** that enable VOs to run workflows and data storage systems across all OSG sites
 - ◆ OSG platform is essential for main stakeholders, the LHC and other large experiments and VOs
 - ★ **OSG Direct facility** uses *job manager overlays* to effectively harvest CPU resources opportunistically from all OSG sites, that stakeholder VOs would otherwise leave idle
 - ◆ used by a large and diverse **community of researchers** and **science platforms**
 - ◆ basis for being an **XD Service Provider**, to which PIs acquire access through XRAC allocations
 - ★ **OSG Connect** platform provides a login and group management service for VOs and individuals with HTC workloads
 - ◆ connect **campus users, Campus Grids** as a service

Open Vision of OSG

Frank Würthwein's talk

- Open Science
- Open Facility
- Open Software Stack
- Open Ecosystem

Miron: Overlay Job Manager and Blueprint

◆ “We do not develop new technologies” was seen as a huge strengths!



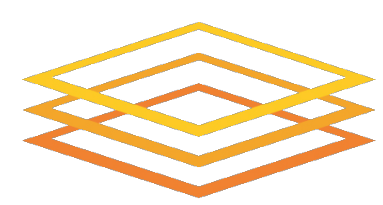
The Technology Gate Keeper Miron's talk

We use the OSG Blueprint (evolving principles) that follows the OSG DHTC (underling) principles to guide our technology roadmap – adoption and promotion.

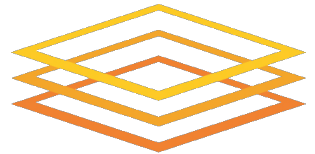
My role is to manage the formulation of the underling principles and the development of the blueprint. What we have today is based on almost 35 years of work in the area of distributed computing and the operation of production computing systems at all scales.

- We do use technologies that do not follow our blueprint.
- We promote and help advance technologies that follow our blueprint.
- We do not develop new technologies.

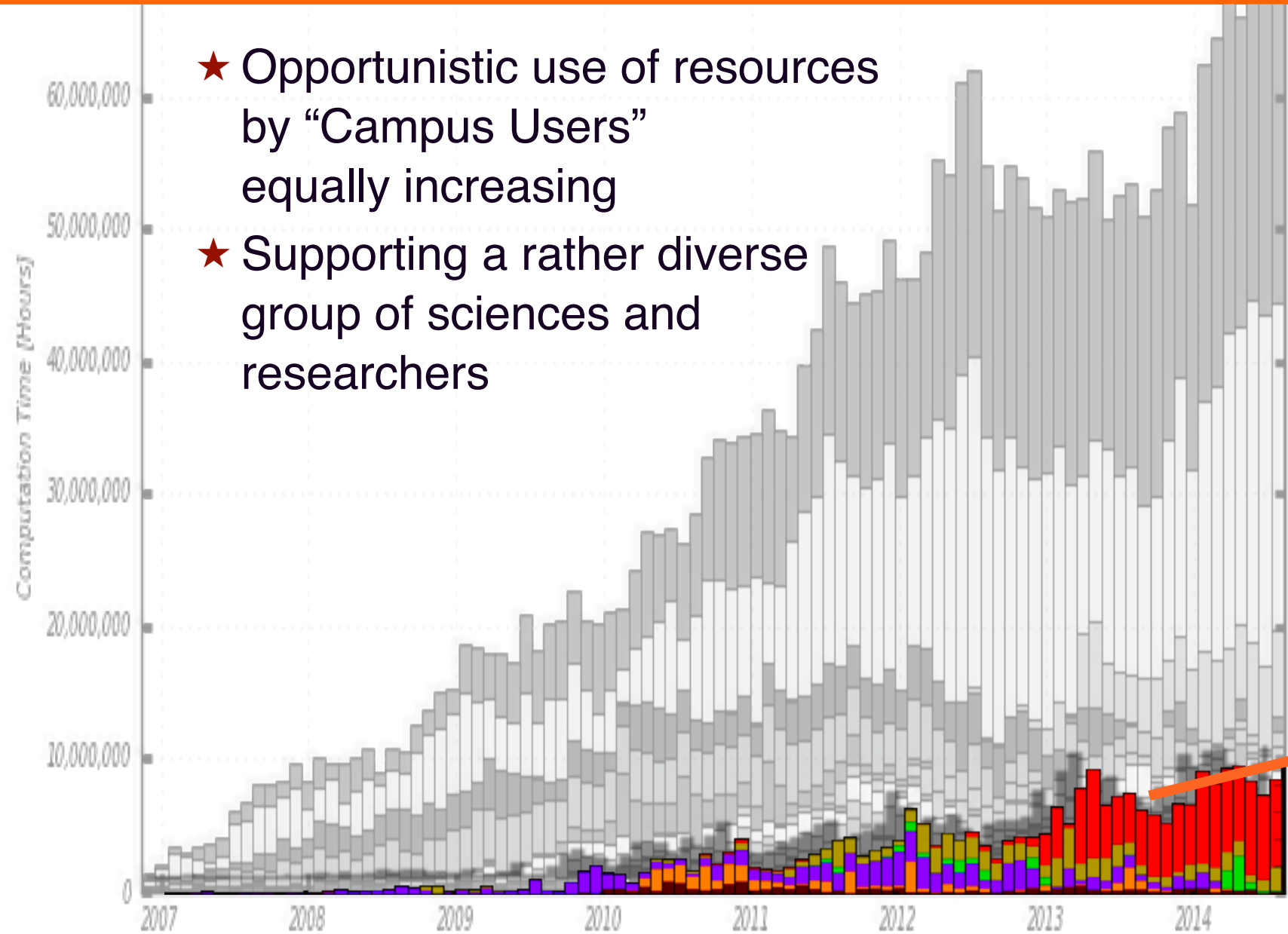
- OSG's Emphasis needs to be on
 - Fast transient resource contribution and provisioning
 - Authentication Management
 - Providing Data and Dynamic Storage
 - Making Network Monitoring Data an Integral Part of the OSG ecosystem
- “OSG Connect” unites most of the above and is the strategic direction of OSG



We put a major emphasis on the value OSG creates by enabling science to profit from opportunistic resources



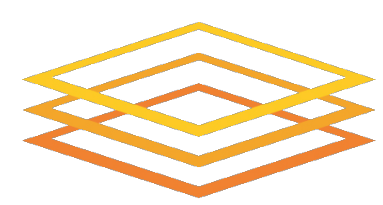
Growing Use of "Owned" and of "Opportunistic" Resources



- ★ Opportunistic use of resources by "Campus Users" equally increasing
- ★ Supporting a rather diverse group of sciences and researchers

Opportunistic Use?

>90M CPU hours opportunistic use past 12 Months



Workplan at the Midpoint of Project: Main Areas OSG Delivers On

◆ Operations

- ★ provide the OSG platform/eco system of DHTC services, sites, software **to enable VOs** to run workflows and data systems **across OSG sites**
 - ◆ running a world-class unique diverse set of services, enabling more than 100 sites
 - ◆ infrastructure services, operations support, cyber security and incident response etc
 - ◆ main customer is the LHC, and other large experiments/VOs

- ★ provide a production quality **HTC facility** built on **harvesting resources opportunistically** from OSG sites, for a large & diverse community of researchers and science platforms
 - ◆ delivering amongst others as a XD Service Provider, through XRAC allocations
 - ◆ delivering to science gateways e.g. for biology or medical applications etc

- ★ provide **other added value:**
 - ◆ user and host certificates (OSG CA, the follow-up of DOEgrids CA)
 - ◆ software distribution services (OASIS based on CVMFS)
 - ◆ network monitoring and dashboard
 - ◆ ...

◆ 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
 - ◆ a small number of focus areas like HTCondor-CE, scalability, authz, etc,
- ★ software “factory”: packaging, system testing, patching

◆ Campus Grids

- ★ OSG Connect service
- ★ Campus Infrastructure Community

Evolving The Workplan For 2nd Half of OSG Project

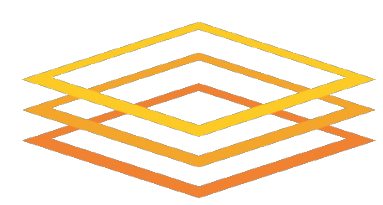
1. Provide value to VOs and resource owners by enabling DHTC and resource sharing for their set of applications
2. Provide opportunities for the Long Tail of Science
3. Promote use of DHTC for science

1. Provide Value to VOs and Resource Owners

- ◆ ...by enabling DHTC and resource sharing for their set of applications
- ◆ Main part of workplan continues to be driven by large VOs, resource owners
 - ★ LHC preparing for Run2 w/ large increase in computing demands in 2015
 - ★ plan for the luminosity upgrade of the LHC (in ~ 2022), will require
 - ◆ computing innovations to help with transition by orders-of-magnitudes scales
 - ◆ new development /deployment approaches to DHTC workflows, distributed data mgmt
 - ★ Again OSG will play important role working with experiments on their upgrade plans, while ensuring production infrastructure remains usable and robust.
- ◆ Intensity Frontier: make IF experiments run effectively on OSG
 - ★ User Support is working to get IF groups running on OSG
 - ◆ at Fermilab NOvA, g-2, MicroBoone, mu2e, and LBNE (35Ton)
 - ◆ also Belle-II and SNO+, to allow opportunistic use of resources
 - ◆ Belle-II and NOvA routinely running MC production, SNO+ has run validation tests, MicroBoone is finalizing code and testing for MC Data Challenge in mid-August, and LBNE has used OSG for fast- simulation ahead of the DOE review last May
- ◆ These stakeholders also directly profit from other OSG offerings
 - ★ OSG-CA, work on security and trust relationships, network monitoring, etc...
 - ★ and, increasingly, opportunistic use of non-owned and/or “temporary” resources

2. Provide Opportunities for the Long Tail of Science

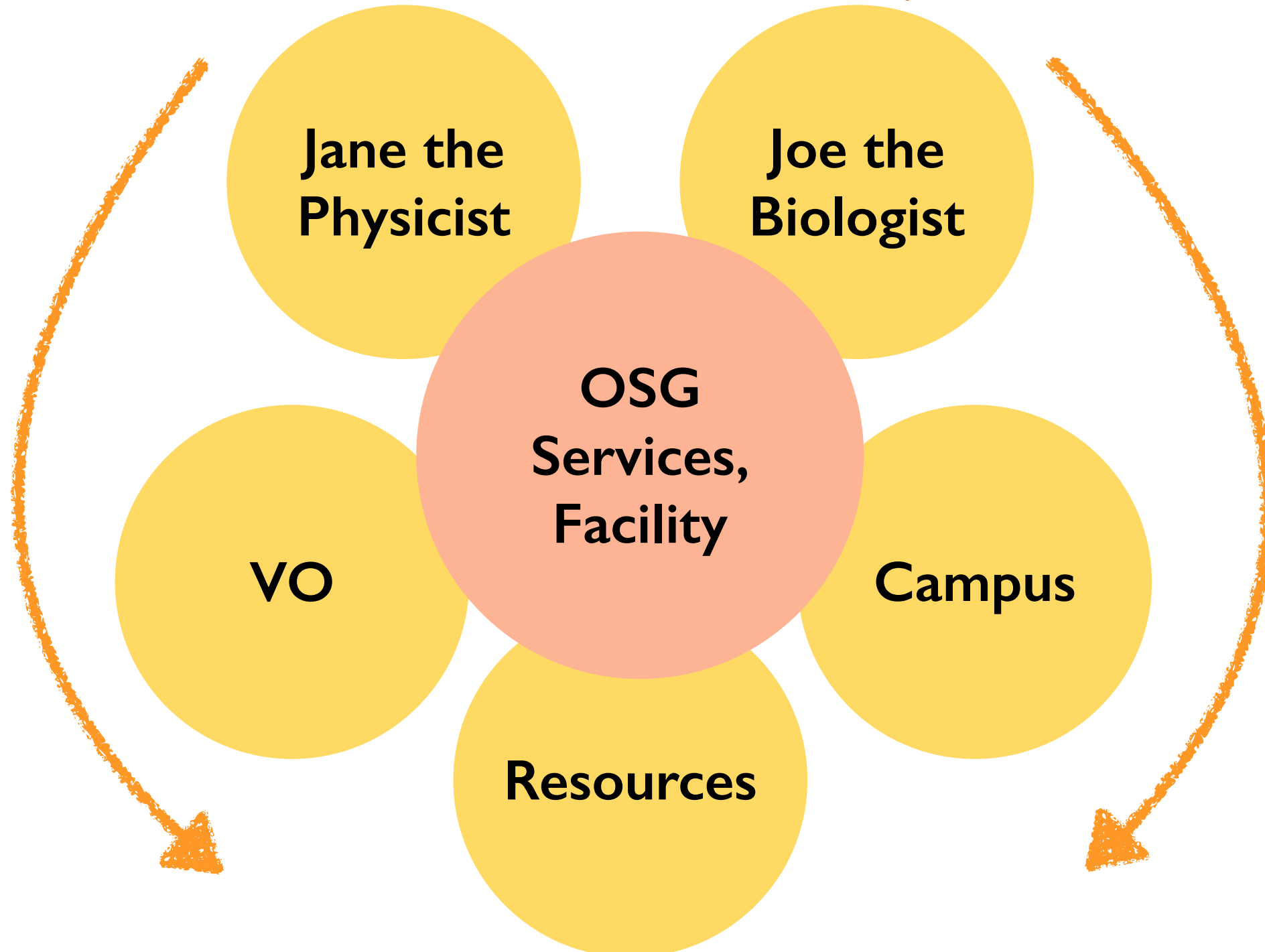
- ◆ Implementing the OSG Open Vision → FWürthwein's talk
 - ★ extend benefits of DHTC to broadest, most diverse set of scientific communities
 - ★ reach broad spectrum of scientists at variety of scales
 - ◆ from individual users at single campus to multi-institutional experiments
- ◆ Needed to diversify support and access models
 - ◆ XD Service Provider (XRAC), OSG Connect (logins), Science Gateways (e.g. Galaxy)
- ◆ Long Tail of Science
 - ◆ researchers at campuses that don't own large resources but profit from DHTC
- ◆ Prevent growth of capability gap between two extremes: LHC ... Long Tail
 - ◆ while “exascale” problems need to be solved for the large collaborations, it is equally necessary to ensure solutions are available for the many scientists challenged at the terascale and petascale on their passage to the exascale
- ◆ Keep up with increasingly dynamic and heterogeneous environments
 - ◆ ensure that domain scientists with limited computing expertise can use them
- ◆ **campuses** to play role filling this gap, supporting researchers in DHTC
 - ◆ for our user community, including LHC, it is crucial we continue to include the campuses!
- ◆ strategic importance to partner w/ projects supporting campuses
 - ◆ like ACI-REF

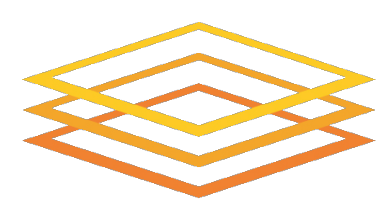


OSG Services and Facility Ecosystem

Open Science Grid

VO-centric Model ⊕ **Campus User Model**

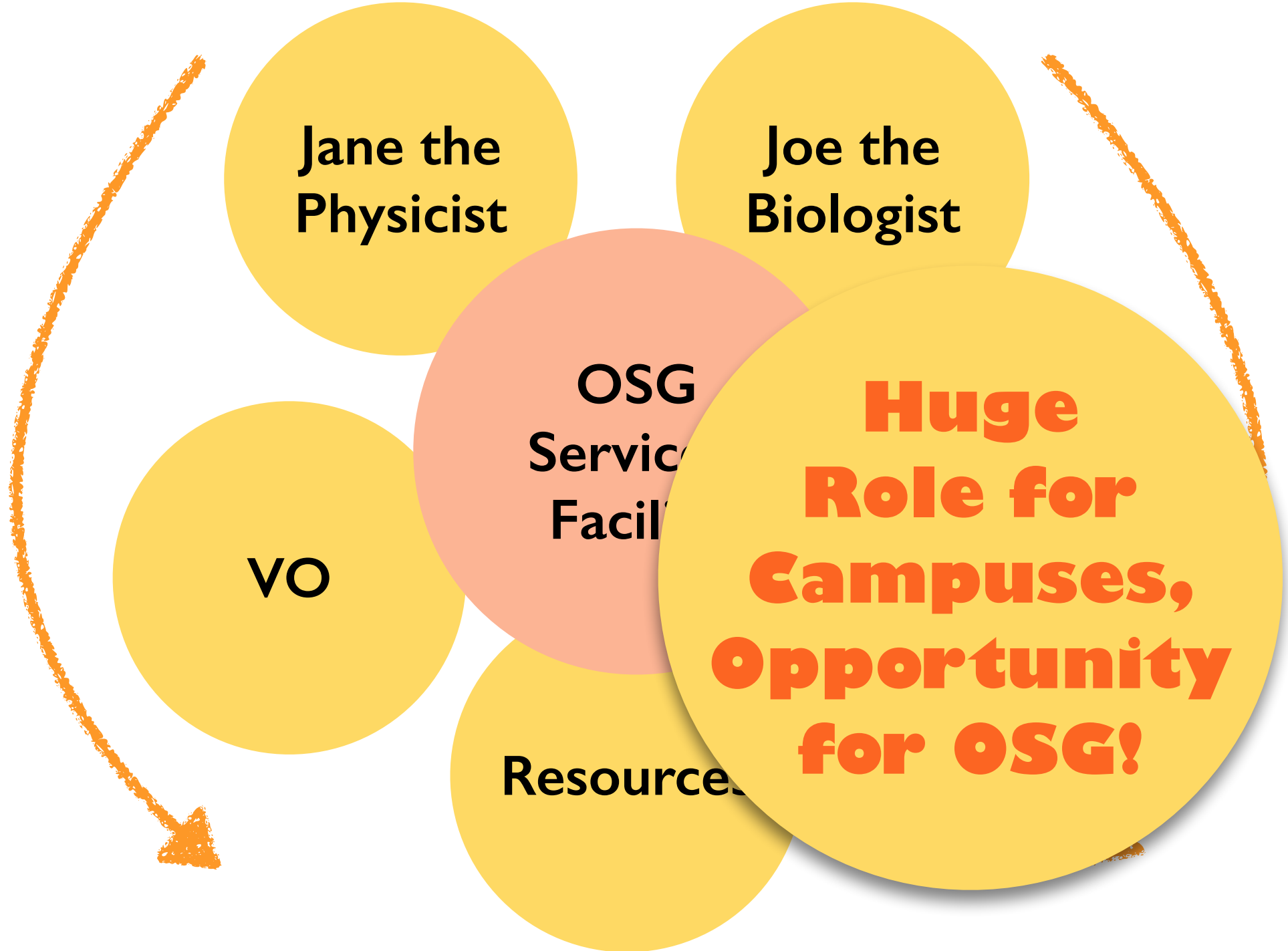


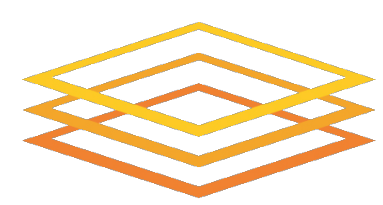


OSG Services and Facility Ecosystem

Open Science Grid

VO-centric Model ⊕ Campus User Model

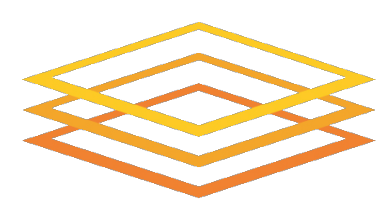




OSG Strategy For Campus

Open Science Grid

- ◆ “Submit local, compute global” using Bosco tool
 - ★ downloadable self-installable, self-configuring robust tool to allow researchers to do HTC on whatever resources are easiest available to her/him
- ◆ New complementary approach: OSG Connect provides “OSG as a service”
 - ★ campuses connect to OSG through set of services hosted & operated by OSG
 - ◆ allows campuses to start using OSG without having to deploy services or configurations
 - ★ “encapsulates” basic infrastructure services
 - ◆ network services (globus.org), basic data services, other services researchers require
 - ◆ identity management services, use existing campus IDs across OSG
 - ★ access to compute resources based on HTCondor technologies
 - ◆ setup easy access to a variety of facilities and resources
- ◆ Provides campuses with the services to connect researchers to OSG
 - ★ allows them to create Campus Grids, or extend applications across campus
 - ★ campuses don’t have to build their own — OSG runs these services for them
- ◆ Initial set of services exists, new campus grids are being established
 - ★ getting some traction also in Atlas and CMS as a solution for Tier-3



Bringing in New Types of Resources

Open Science Grid

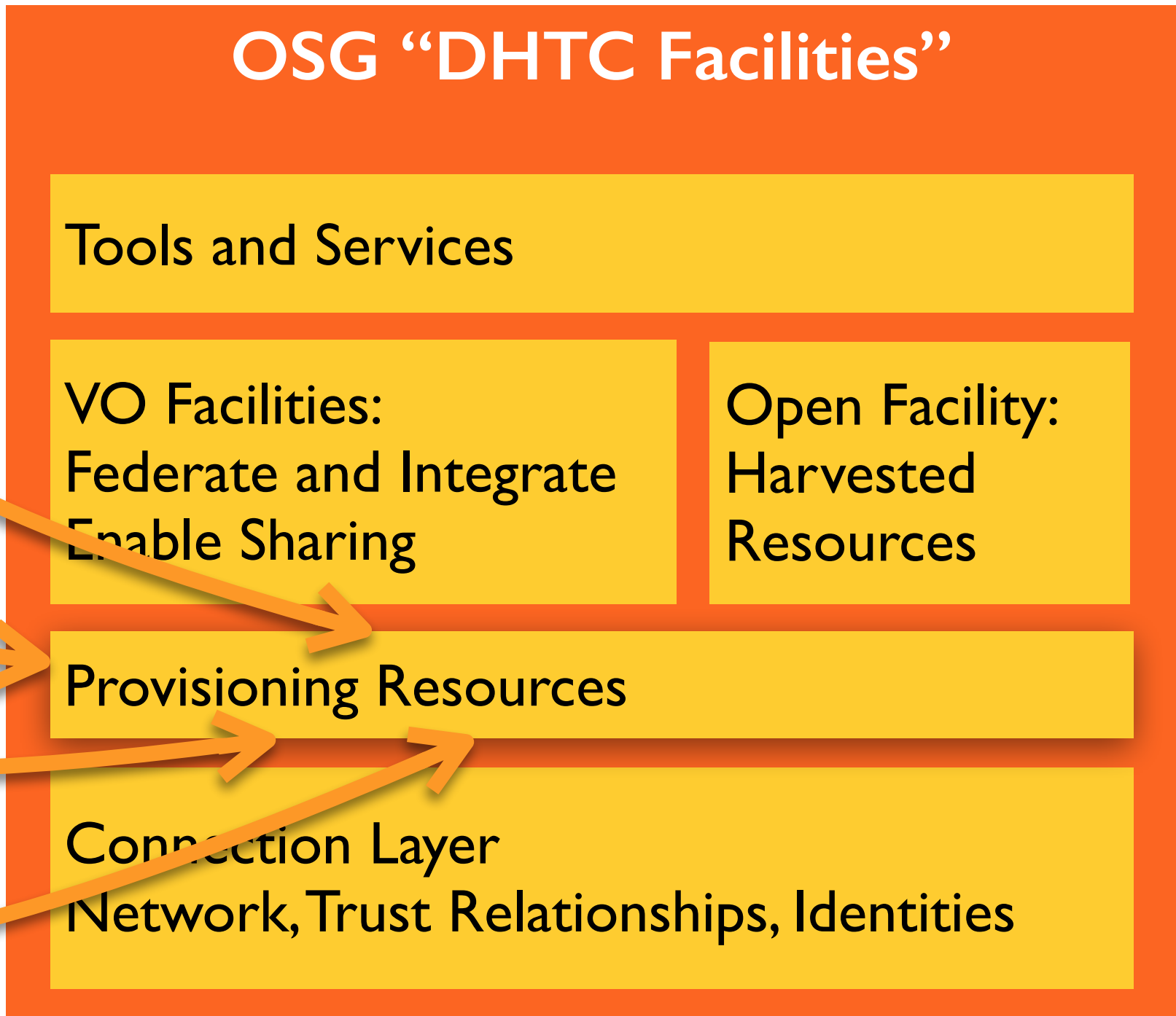
Processing

VO-owned

Across the campus

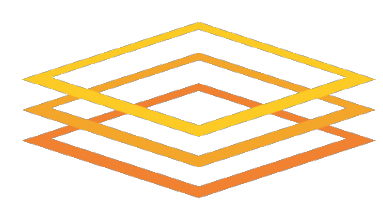
Allocation at HPC

Industry Commodity Services
Google, Vodaphone



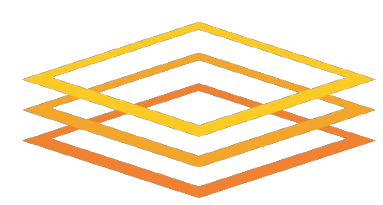
◆ Focus on **Dynamic Resource Provisioning**

★ Statically federated resources need to be integrated with dynamically allocated resources causing new challenges for resource planning, acquisition, provisioning

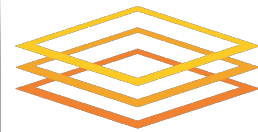


3. Promote use of DHTC for Science

- ◆ HTC is a paradigm used and explored broadly throughout DOE and NSF science
 - ★ including genomics, structural biology, high energy physics, nuclear physics, etc
- ◆ Effective DHTC requires making use of a **diverse set of resources** to achieve maximum capacity and cost effectiveness, and **establishing mutual trust** among the participants while still respecting their **autonomy**
 - ★ Achieving these DHTC attributes presents **significant research and technical challenges** as the applications and technologies increase in size and diversity
- ◆ OSG contributes intellectual resources, expertise, and solutions needed to **advance the computing throughput of DOE and NSF science**
 - ★ we have a number of important activities in this area that are crucial to further develop adoption and use of DHTC across OSG stakeholders
 - ◆ Focus Groups, provide technical leadership/guidance LHC, IF, others
 - ◆ Campus Infrastructure Communities, workshops etc
 - ◆ Tutorials and documentation
 - ◆ Evolve partnership with XD/XSEDE, WLCG, large experiment S&C organizations
 - ◆ Establish partnership with ACI-REFs
 - ◆ OSG User School
 - ◆ Publish OSG Research Highlight monthly

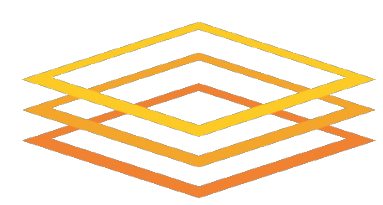


- ◆ The committee wanted to hear more about the LHC, how the OSG is still relevant to the LHC etc — which we had not focused the presentations on
- ◆ we provided sound bytes (which was what they were asking for mostly) —



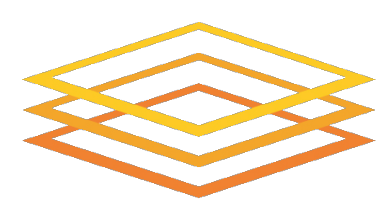
LHC Continues to Depend Critically on OSG

- ◆ OSG enabled the US LHC to deliver the pledged computing resources
 - ★ for CMS, the US delivered more than 40% of the Tier-1 resources worldwide and 35% of the physics data analysis capacity
 - ★ over the past year, US provided more than 0.5 Billion CPU hours to WLCG
- ◆ OSG enabled USLHC scientists to be leaders and active participants in the discovery process, by delivering data to the Tier-2 and Tier-3 center
 - ★ for CMS > 5 PB to Tier-2 and > 1 PB to Tier-3 centers over the past year
- ◆ OSG enabled US LHC to contribute more than 2.5B simulated events to CMS over the past year (Atlas ?)
- ◆ OSG enabled US LHC to provide first-in-class Tier-1 and Tier-2 facilities
 - ★ measured in availability and reliability metrics that OSG collected and reported to WLCG
 - ★ 6 of the 7 USCMS Tier-2 sites were in the top-10 (of 52 worldwide) in terms of availability and reliability, demonstrating the reliability of OSG services
- ◆ OSG ran and processed more than 6,000 probes per day to generate reports to WLCG
- ◆ OSG is positioned to support the Run 2 data challenges in the same way we supported the data challenges Run 1



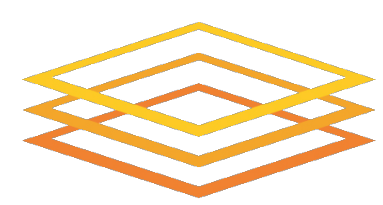
The OSG Program of Work Needs to Continue!

- ◆ Required by LHC and IF experiments, and other large VOs
- ◆ Sites and VOs, Campuses would have to replace services
 - ★ PKI/IdentityManagement/CA, Network Monitoring, Information Systems, etc etc



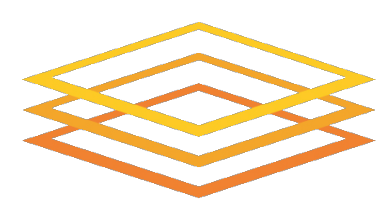
The OSG Consortium and its Eco-System will Continue

- ◆ Consortium expressed it's desire to sustain a long-term collaboration
 - ★ All participants profit from and value the consortium
- ◆ Initial discussions on “sustainability”, including thinking about “extending the business model” last year
 - ★ was seen as too early in the lifecycle of the current project, but will be picked up again



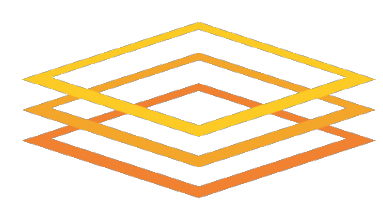
The OSG project currently seen as the best possible organizational form to continue this work and effort

- ◆ The OSG Project is a very well established way to work together, to move forward, to react to and fulfill stakeholder needs
- ◆ for HEP, the P5 recommendations were explicit to endorse cross-experiment approaches like the OSG
- ◆ VOs and sites expressed that they are looking for technical leadership
- ◆ it remains important to retain and further develop a work force that will drive, guide and grow OSG



The OSG project produces value every day

- ◆ As an example, the opportunistic facility creates “Computational Value” out of un-utilized resources
 - ★ at a value of ~\$5M/year assuming CPU to cost ~5 cent/h elsewhere

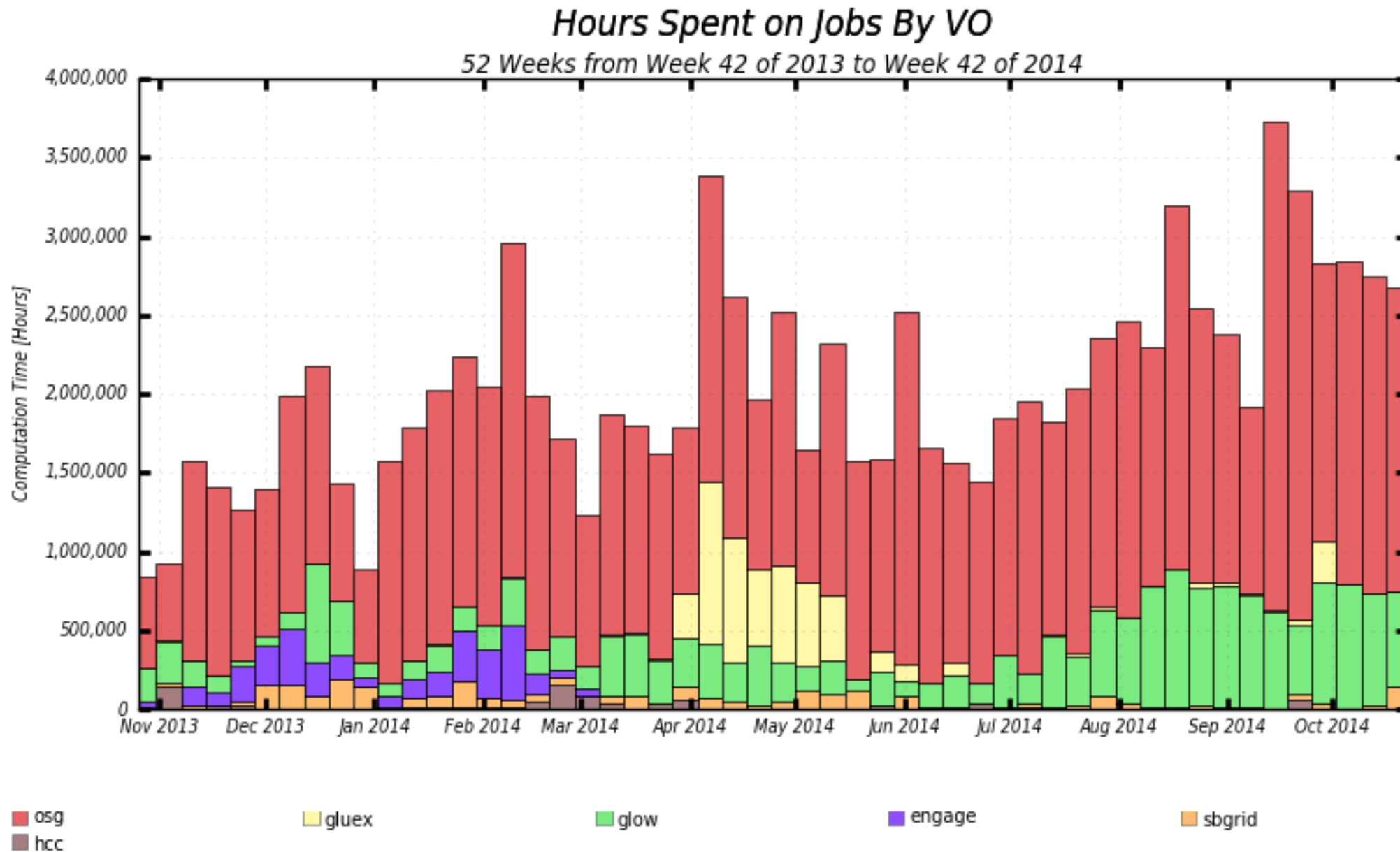


OSG as an Open Facility is still growing

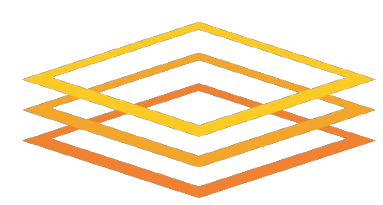
Open Science Grid

◆ 106M CPU Hours delivered opportunistically over 12 months!

★ In the last 10 weeks, we harvested 28M (annual rate of 140M) ==> should exceed 120M for OSG year3 (compared to 89M for year2 -- 30% growth)

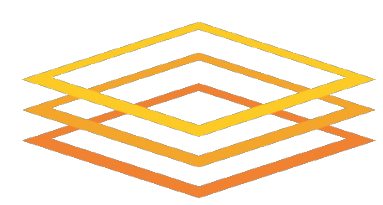


Maximum: 3,723,196 Hours, Minimum: 837,511 Hours, Average: 2,042,376 Hours, Current: 2,675,398 Hours



It's not too early to start preparing for continuation beyond year 5

Encourage the Council to fully engage



Review Committee Questions

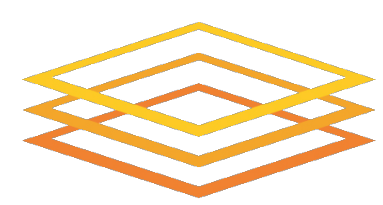
Open Science Grid

- ◆ Does the LHC continue to depend critically on OSG? Measures of success? Have you met them? Readiness for Run2?
 - ★ prepared a dedicated slide, to give soundbites
- ◆ Metrics regarding operational excellence
 - ★ drilled into ticket response/resolution, we will update our metric here
- ◆ Project management: explain how you prioritize, track milestones
 - ★ led them through the process, including stakeholders (council) and ET
- ◆ Interactions with other grids?
 - ★ showed how well we're connected, overlap vs complement

General comments from the review committee

◆ Generally very positive and congratulatory

- ★ OSG provides a high level of availability, maintaining levels of service that seem to be among the best for grid providers and approach that of commercial (cloud) providers
- ★ The panel commends OSG on its success in serving the HEP and NP communities, and on the development of OSGConnect, which lowers the barriers to use for both individual PI's and the larger VO's.
- ★ Striking the right balance between delivering on critical, immediate LHC Run2 needs and outreach beyond PY5 will require constant vigilance. Don't let the long tail of science wag the dog
- ★ Extension of OSG resources to eg. XSEDE, NERSC, and Clouds using OSG Connect holds promise of access to new capable resources
- ★ Scalability of OSG by x2 (e.g. for LHC Run2) seems achievable, but must be demonstrated soon



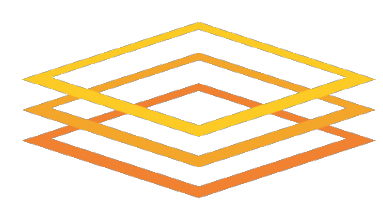
A small number of Recommendations

Open Science Grid

- ◆ Evaluate the goals and program of work for years 4 and 5, including their appropriateness and value to the communities served.
 - ★ 1. If outreach to non-LHC/non-HEP communities is to be a focus of PY4-5 and beyond, a funding model involving these communities needs to be developed.
 - ★ 2. The management & decision making process in OSG needs to be more clearly articulated and demonstrated. A transparent record of actions proposed and plans defined with understandable rationale and success metrics should be available to Program Offices following the yearly Executive Board Meeting.
- ◆ Discuss the appropriateness and effectiveness of plans for collaboration, personnel, and management organization, and for meeting demands on computing and data processing...
 - ★ 1. For the years 4 and 5 sharpen the articulation of each of the objectives.
 - ★ 2. Improve the reporting of the operations metrics to justified that the appropriate demands are being met

◆ Comments

- ★ OSG has few/no competitors in the US
- ★ commends OSG's focus on using externally developed software and limiting their scope to hardening and integrating software into an operational capability, which gives the project significant agility
- ★ OSG has a unique organizational model focusing on user communities (versus grids) as stakeholders
- ★ At present, the support costs for individual researchers (e.g. in terms of trouble tickets) does not appear to large relative to the total, indicating that supporting individuals is not drawing resources away from supporting core customers.
- ★ OSG team made a compelling argument that developments created to help individual researchers have also benefitted core customers
- ★ should continue to opportunistically engage with individual researchers and seek to expand the range of disciplines benefitting from its work, but it should also closely track the actual and opportunity costs of this work



We believe the review was a great success, allows us to continue for the 2nd half of the project, and puts us on a good track for a continuation of OSG beyond project year 5!

Thanks to all who helped and to a great OSG Team!