Open Science Grid

Outline of the Workplan
for the 2nd half of the OSG Project
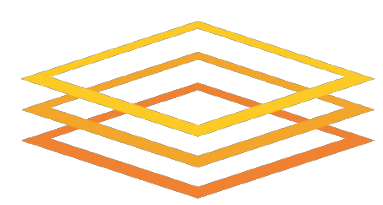
OSG Agency Review
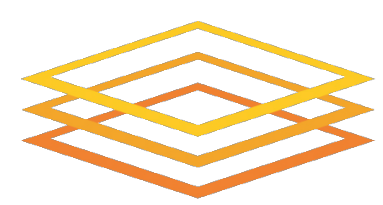**Arlington, VA, August 19, 2014**

LATBauerdick
OSG Executive Director

Overview

- ✦ This review is of the workplan for the 2 out-years of the OSG project
  - ★ to meet the growing and changing computing needs of U.S. LHC program
  - ★ to 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
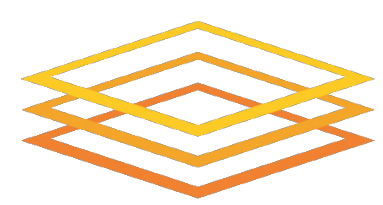   - ✦ connect **campus users**, **Campus Grids** as a service

OSG in Numbers

✦ OSG Delivers up to 2.5 Million CPU hours every day

   ★ almost 750M hours of Distributed High Throughput Computing per year, of which **>~ 90M were provided as "opportunistic resources"**

   ★ about 60% go to LHC, 20% to other HEP, **20% to many other sciences**

✦ OSG has a footprint on ~120 campuses and labs in the U.S.

✦ OSG transfers ~1 PetaByte of data every day

✦ Supports active community of **20+ multi-disciplinary research groups**

OSG Project Facts

✦ OSG Project is reaching the half-time point of its current 5 year run

★ agency review in August of our work plan for 2nd half of the project

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

Area Manpower	FTE
Technology	7.8
Release Mgmt	1.5
Operations	8.8
Campus Grids	2.8
Networking	0.3
Security	2.1
User Support	3.0
Project Office & Communications	1.2
Total	**27.5**

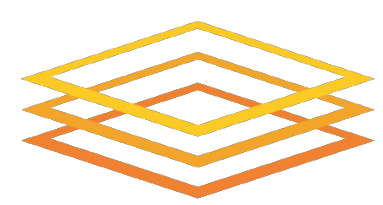| OSG Leadership	Responsibility
Lothar Bauerdick (Fermilab)	Executive Director
Miron Livny (Wisconsin)	Technical Director and PI
Chander Sehgal (Fermilab)	Project Manager, User Support Lead
Ruth Pordes (Fermilab)	Council Chair
Rob Quick (Indiana)	Operations Lead
Michael Ernst (BNL)	ATLAS planning
Frank Würthwein (UCSD)	CMS planning, Resource Manager
Brian Bockelman (Nebraska)	Technologies Lead
Rob Gardner (Chicago)	Campus Grids Lead
Mine Altunay (Fermilab)	Security Officer
Tim Cartwright (Wisconsin)	Software Lead
Tim Theisen (Wisconsin)	Software Releases
Shawn McKee (Michigan)	Network Area Lead

OSG "Size" is Increasing in Terms of CPU Resources

★ Since 2007, 3.1 Billion CPU hours delivered, more than a billion jobs run!

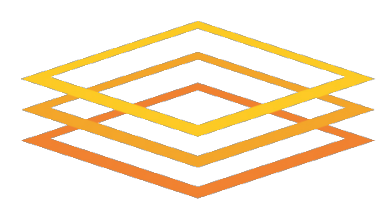★ LHC plans increase in CPU resources at U.S. Tier-1 and Tier-2 over coming 2 years

LHC Run2?

750 M CPU hours past 12 months

Hours Spent on Jobs By VO
405 Weeks from Week 46 of 2006 to Week 32 of 2014

Virtual Organizations

atlas	cms	osg	dzero	cdf
ligo	glow	gluex	dosar	engage
hcc	gridunesp	Other	mu2e	alice
minos	minerva	mars	sbgrid	nova
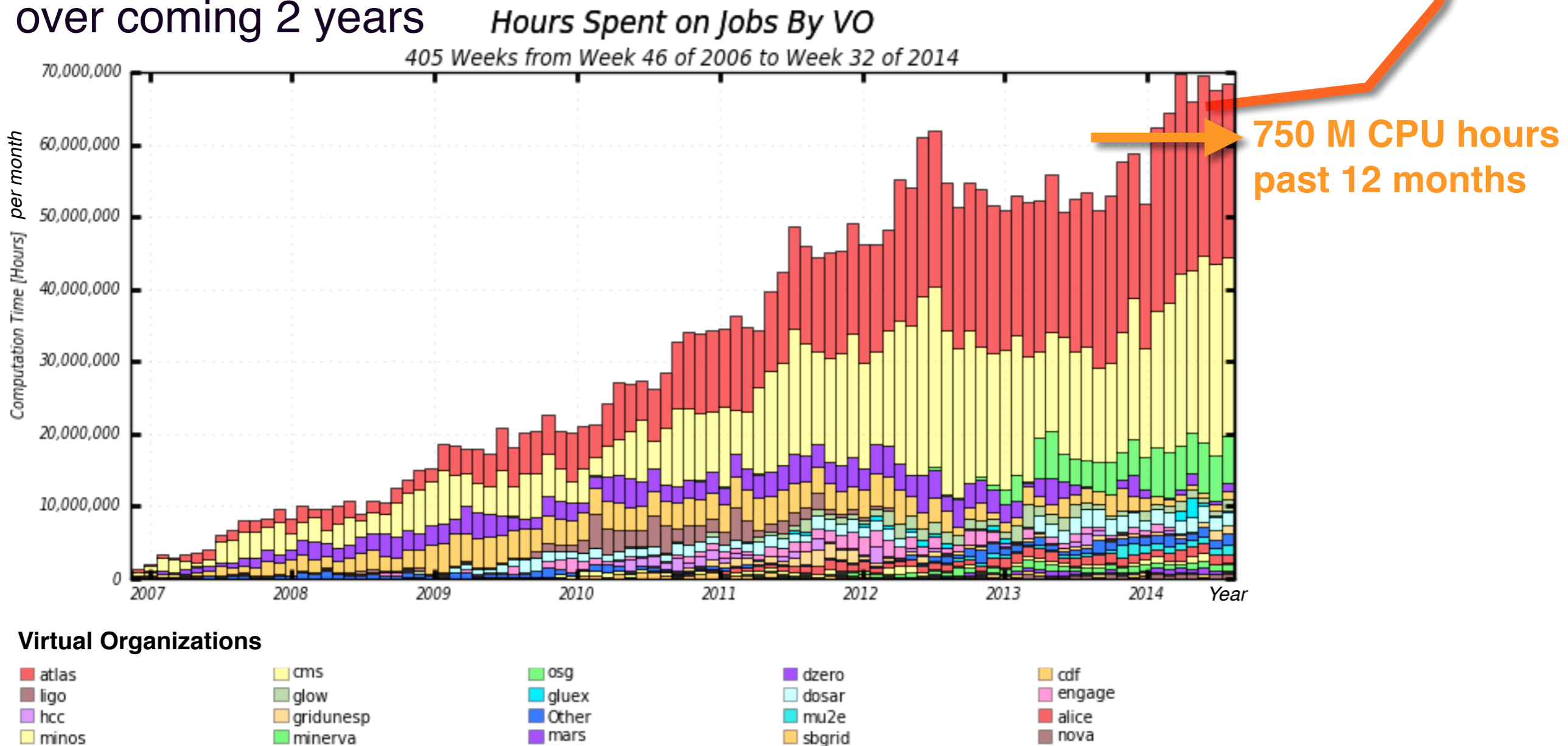
Maximum: 69,871,768 Hours, Minimum: 294,360 Hours, Average: 32,252,999 Hours, Current: 10,916,355 Hours

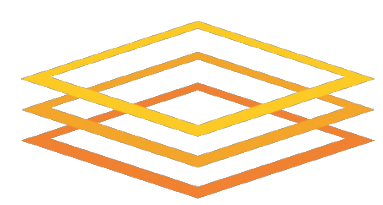# Growing Use of "Owned" and of "Opportunistic" Resources
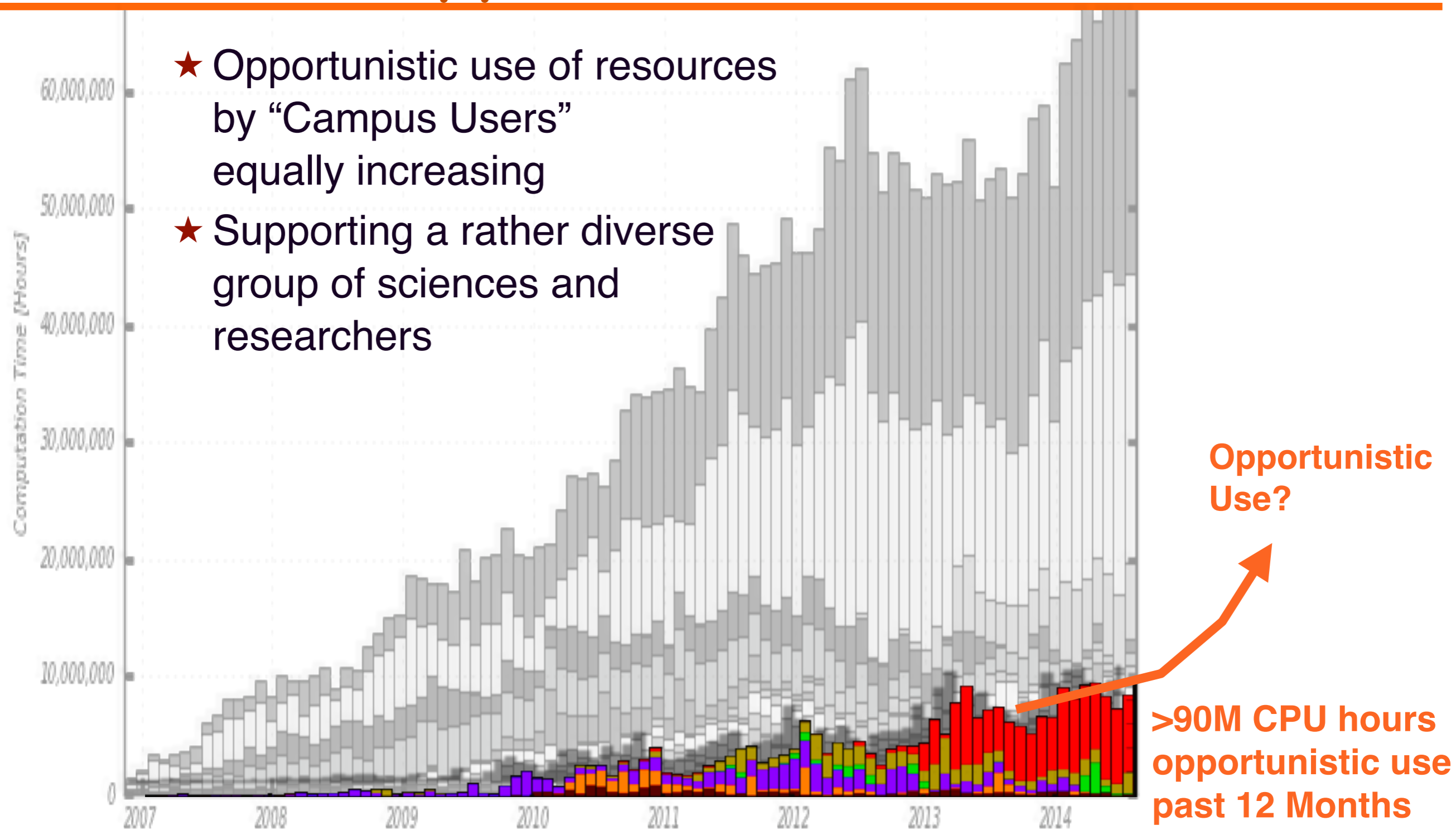
★ Opportunistic use of resources by "Campus Users" equally increasing

★ Supporting a rather diverse group of sciences and researchers

Opportunistic Use?

>90M CPU hours opportunistic use past 12 Months

OSG "Open Facility"

✦ A Distributed High-Throughput Computing Facility based on harvesting otherwise idle resources

- ★ **job manager overlays** using capabilities of HTCondor and GlideinWMS
- ★ effectively harvest CPU resources opportunistically from all OSG sites

✦ Available to a large and diverse community of researchers

- ★ backend for specialized science gateways e.g. for biology and medical applications using web portals such as Galaxy
- ★ is basis for OSG as an Level-2 XD Service Provider, access for PIs through XRAC allocations

OSG Virtual Cluster

OSG DHTC Fabric >100 sites
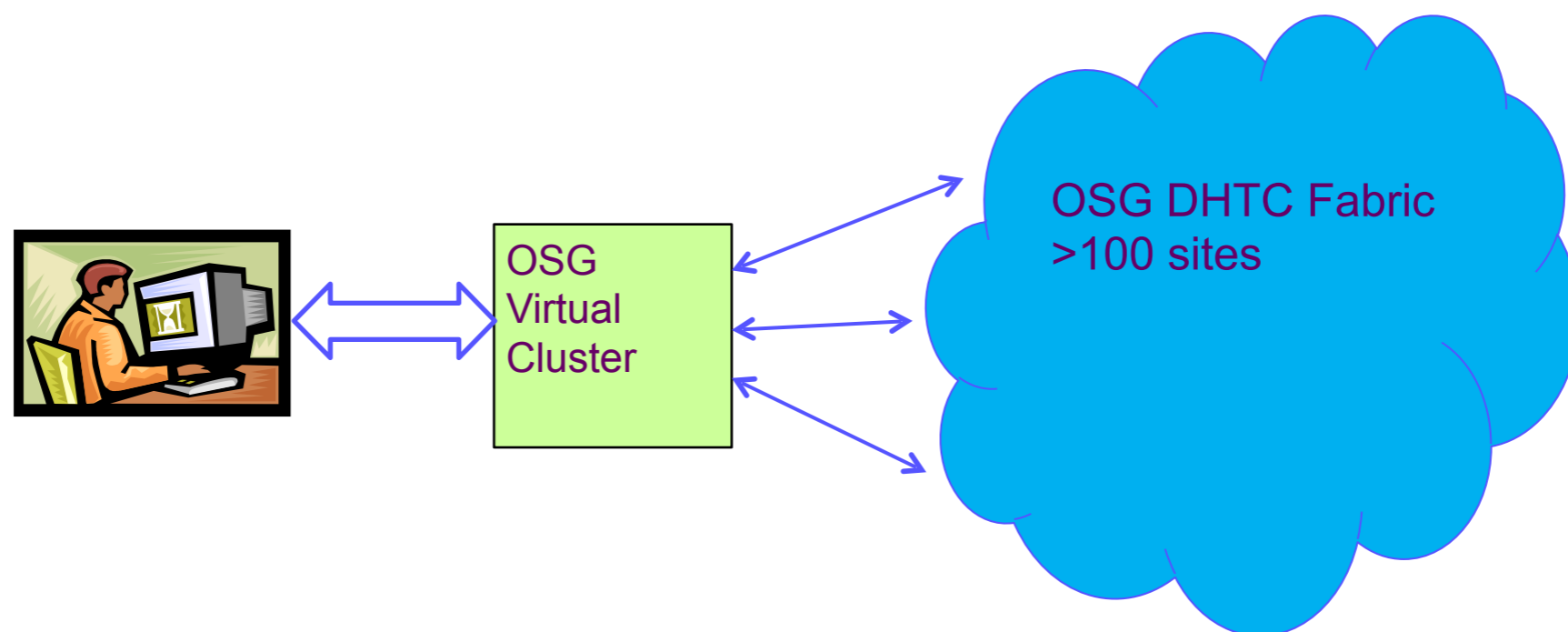
Use of OSG Open Facility

- ✦ OSG VO:
  - ★ XD Service Provider
    - ✦ about 30M hours
  - ★ a number of science applications
  - ★ OSG Connect users ("retail")
  - ★ individual science groups, who also come in through the engage VO
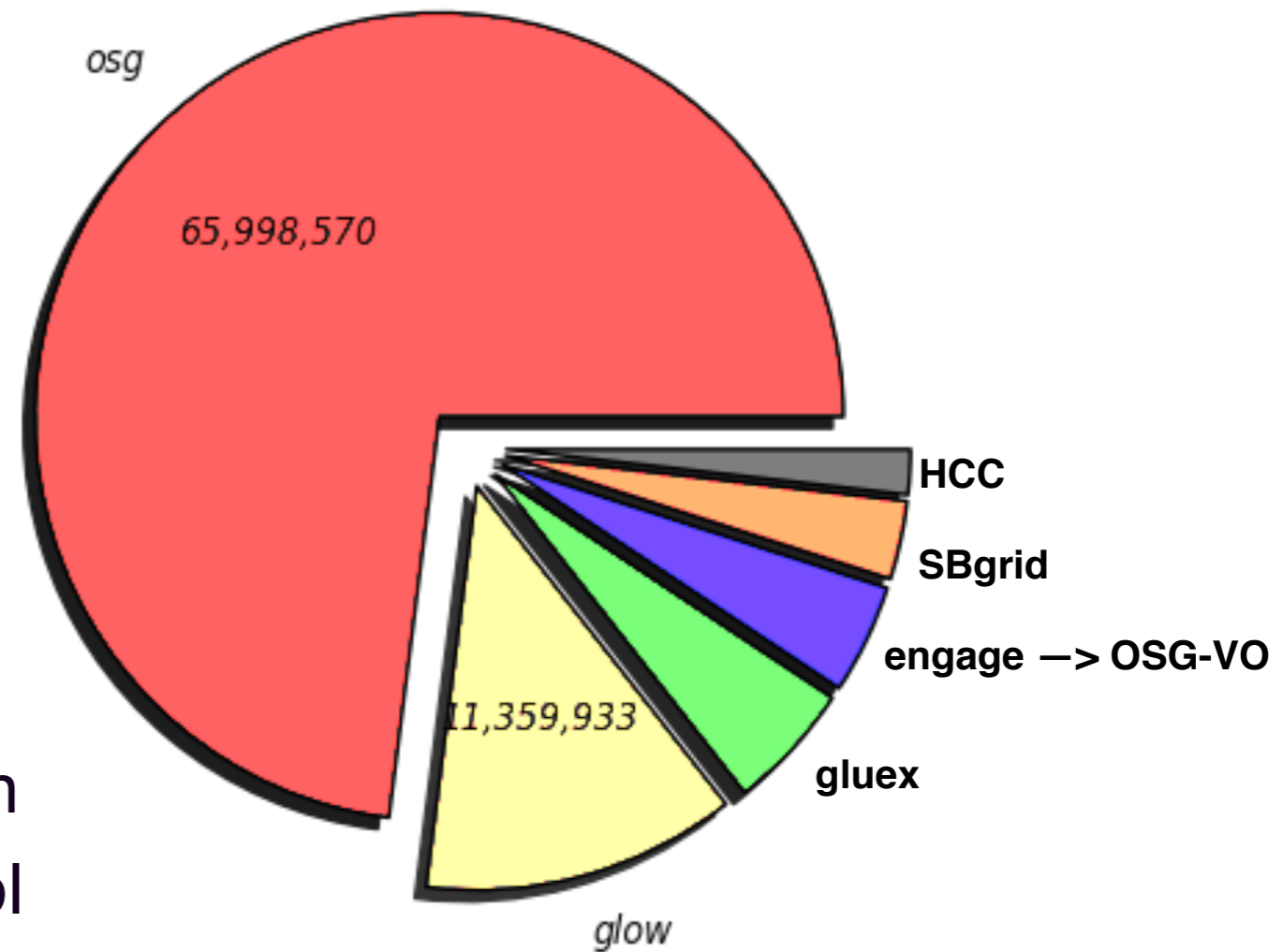- ✦ Users flowing into OSG from campus grids:
  - ★ Glow campus grid at U.Wisconsin
  - ★ SBgrid at Harvard Medical School
  - ★ HCC campus grid at Omaha/Lincoln

Wall Hours by VO (Sum: 90,401,909 Hours)
52 Weeks from Week 33 of 2013 to Week 32 of 2014

- osg — 65,998,570
- HCC
- SBgrid
- engage —> OSG-VO
- gluex
- glow — 11,359,933

Legend:
- osg (65,998,570)
- glow (11,359,934)
- gluex (4,607,346)
- engage (3,967,812)
- sbgrid (2,8...
- hcc (1,621,016)

Wall Hours by VO (Sum: 46,928,493 Hours)
53 Weeks from Week 00 of 2013 to Week 52 of 2013

Medical, Trauma Study — 22,029,113 — D.Krieger/U.Pittsburg, Integrative Biology and Neuroscience

SNOWMASS — 9,972,492 — **High Energy Physics**

SPLINTER Medicine

Duke Nuclear Physics

HEP ECFA Study

Theory HEP

Micro Biology

Computer Science

Legend			
TG-IBN130001 (22,029,113)	SNOWMASS (9,972,493)	SPLINTER (4,174,511)	DUKE-QGP (2,275,491)
ECFA (1,744,646)	TG-PHY110015 (1,004,429)	UMICH (925,155)	RIT (836,827)
Other (797,936)	TG-PHY120014 (510,621)	TG-TRA100004 (444,375)	DETECTORDESIGN (421,086)
EIC (354,954)	UPRRP-MR (274,928)	TG-MCB100109 (262,308)	NESCENT (220,547)
TG-DMR130036 (212,059)	IU-GALAXY (199,146)	KNOWLEDGESYS (164,731)	DUKE (103,137)

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
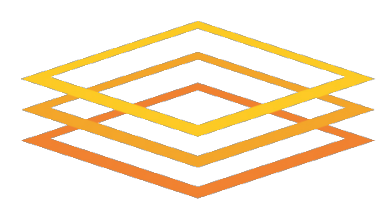
Evolving The Workplan For 2nd Half of OSG Project

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

✦ **…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
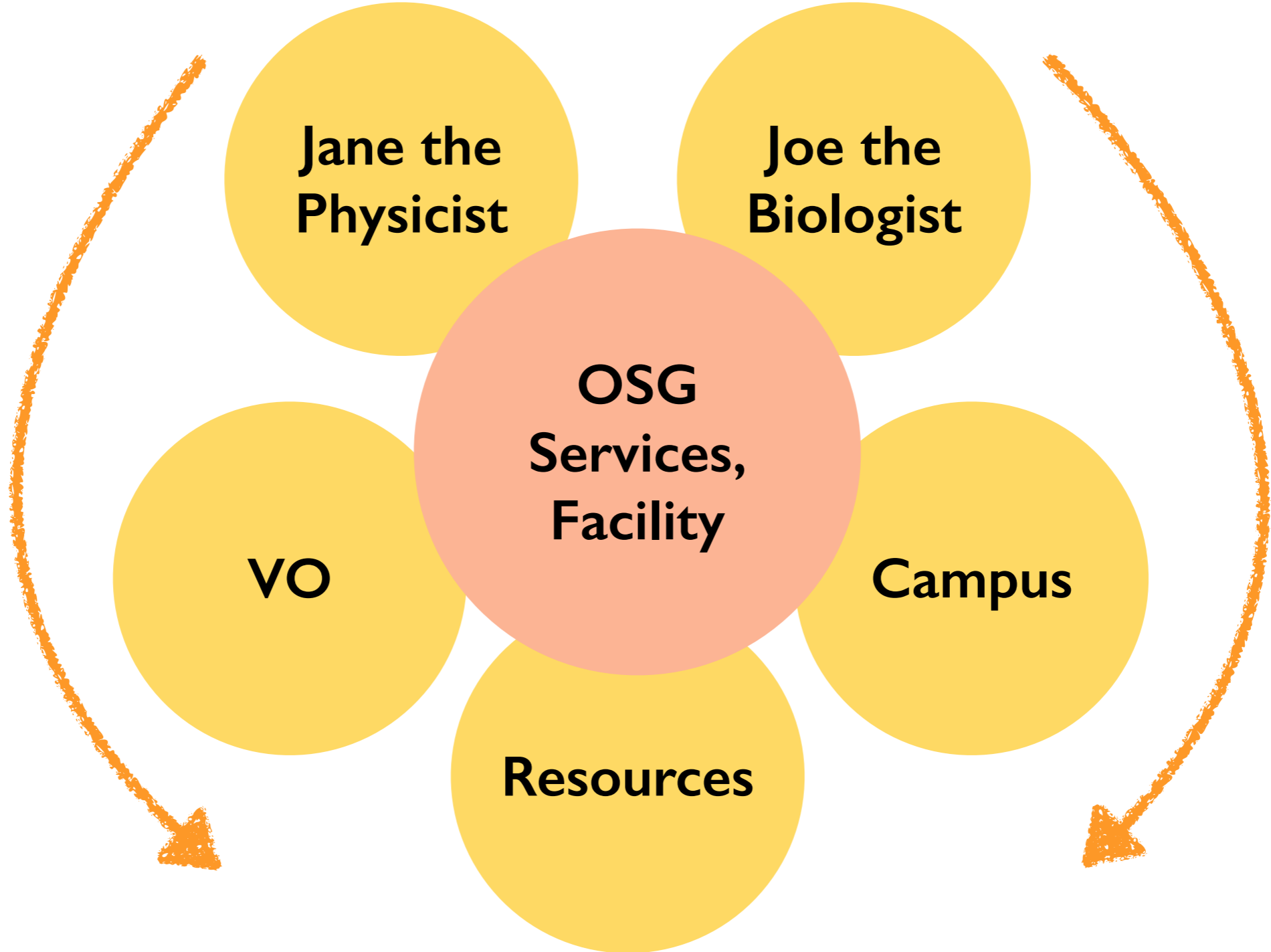
2. Provide Opportunities for the Long Tail of Science

✦ 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!

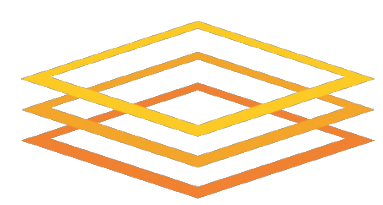✦ strategic importance to partner w/ projects supporting campuses
- ✦ like ACI-REF

OSG Services and Facility Ecosystem
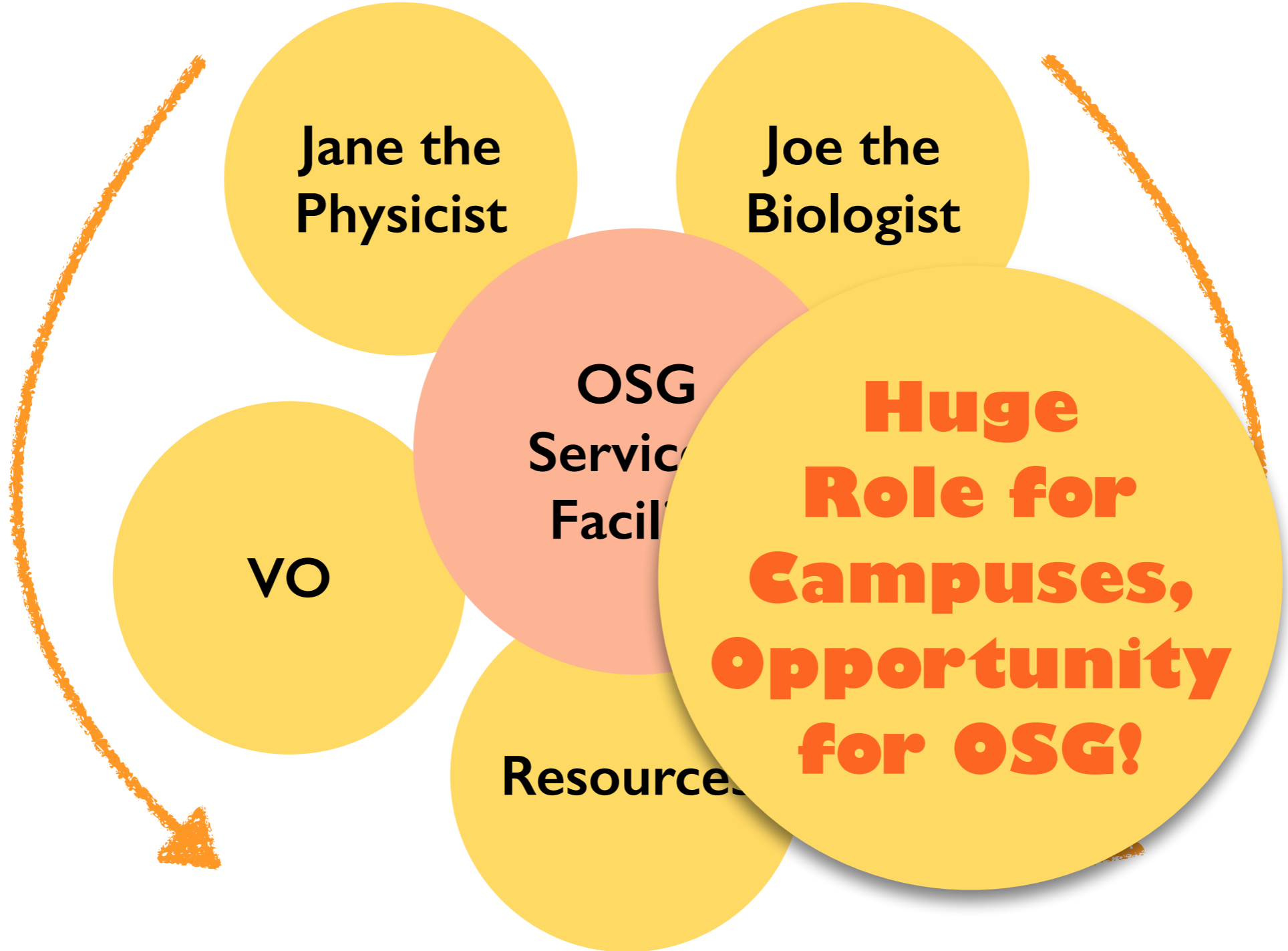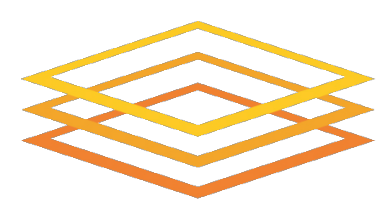
VO-centric Model ⊕ **Campus User Model**

Jane the Physicist

Joe the Biologist

OSG Services, Facility

VO

Campus

Resources

OSG Services and Facility Ecosystem

Open Science Grid

VO-centric Model $\oplus$ **Campus User Model**

Jane the Physicist

Joe the Biologist

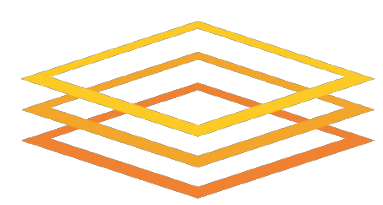OSG Services Facility

VO

Resources

Huge Role for Campuses, Opportunity for OSG!

OSG Strategy For Campus

- ✦ "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
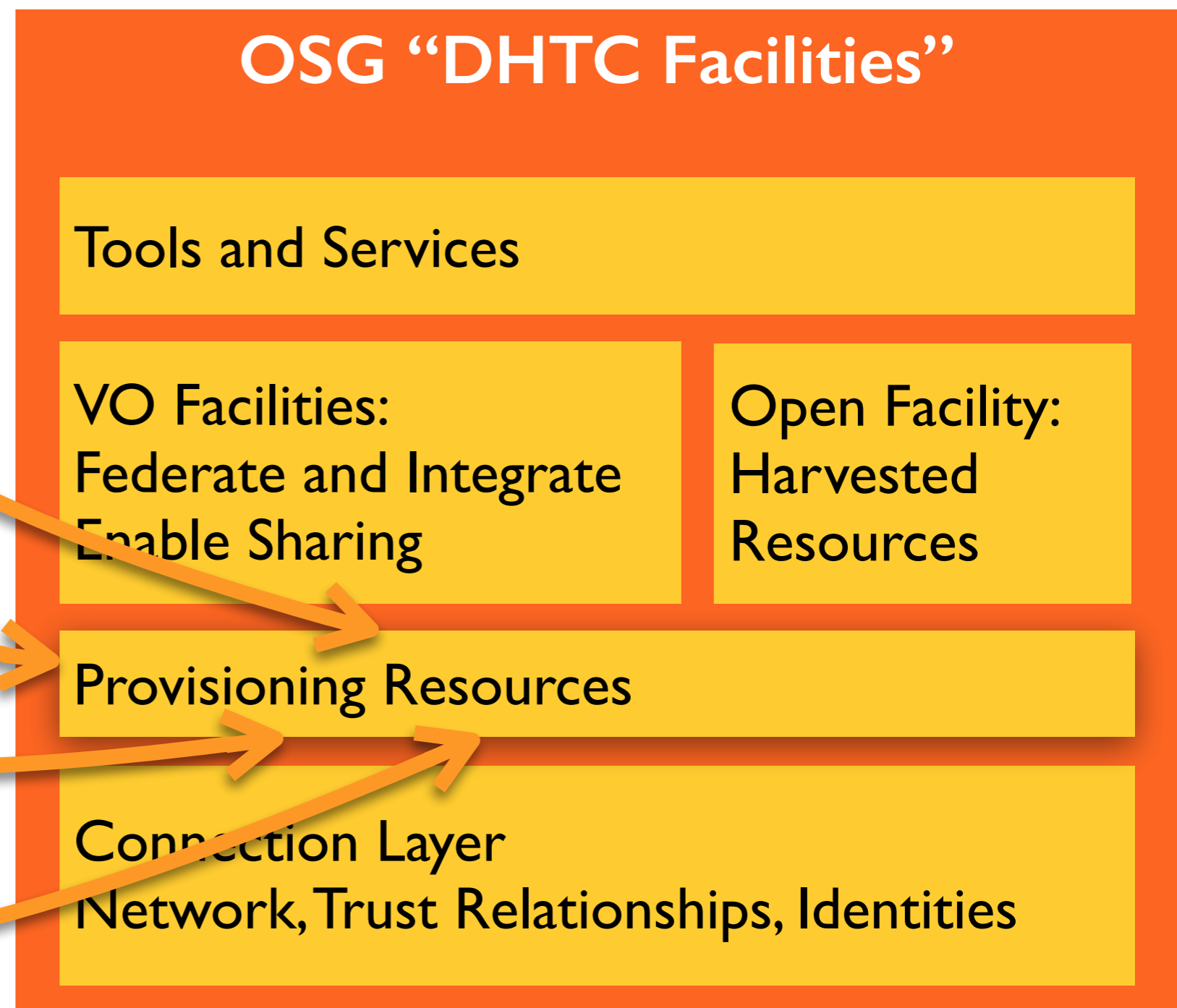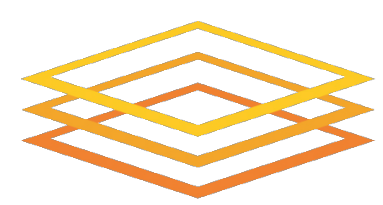
Processing

VO-owned

Across the campus

Allocation at HPC

Industry Commodity Services Google, Vodaphone

OSG "DHTC Facilities"

Tools and Services

VO Facilities: Federate and Integrate Enable Sharing

Open Facility: Harvested Resources

Provisioning Resources

Connection Layer
Network, Trust Relationships, Identities
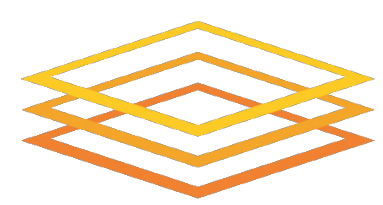
✦ Focus on **Dynamic Resource Provisioning**

★ 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

Summary and Conclusions

- ✦ OSG is doing very well at the half-point of the project
  - ★ smooth operations of a unique distributed set of services and software
    - ✦ successful to sustain and consolidate services and software stack
  - ★ extended service offerings adding important new areas
    - ✦ network monitoring, OSG-CA, a more flexible trust and security model, etc
  - ★ OSG Open Facility is a huge success and produces concrete value
    - ✦ to the NSF XD program, to campuses, to researchers, to portals, etc
  - ★ added innovative ways for researchers and campuses to connect to OSG
    - ✦ OSG connect lowering the barriers of entry to DHTC, adds "OSG As A Service" offering
  - ★ extended customer base, including Intensity Frontier, XD, science portals
- ✦ We have a forward looking work plan for the 2nd half of the project
  - ★ continue to provide value to VOs and resource owners by enabling DHTC and resource sharing for their set of applications
  - ★ a new focus on providing opportunities for the Long Tail of Science
  - ★ promote use of DHTC for science
  - ★ strengthen our important partnerships with WLCG, XSEDE, ACI-REF, …
- ✦ We look forward to continue beyond year-5 of the OSG project!