



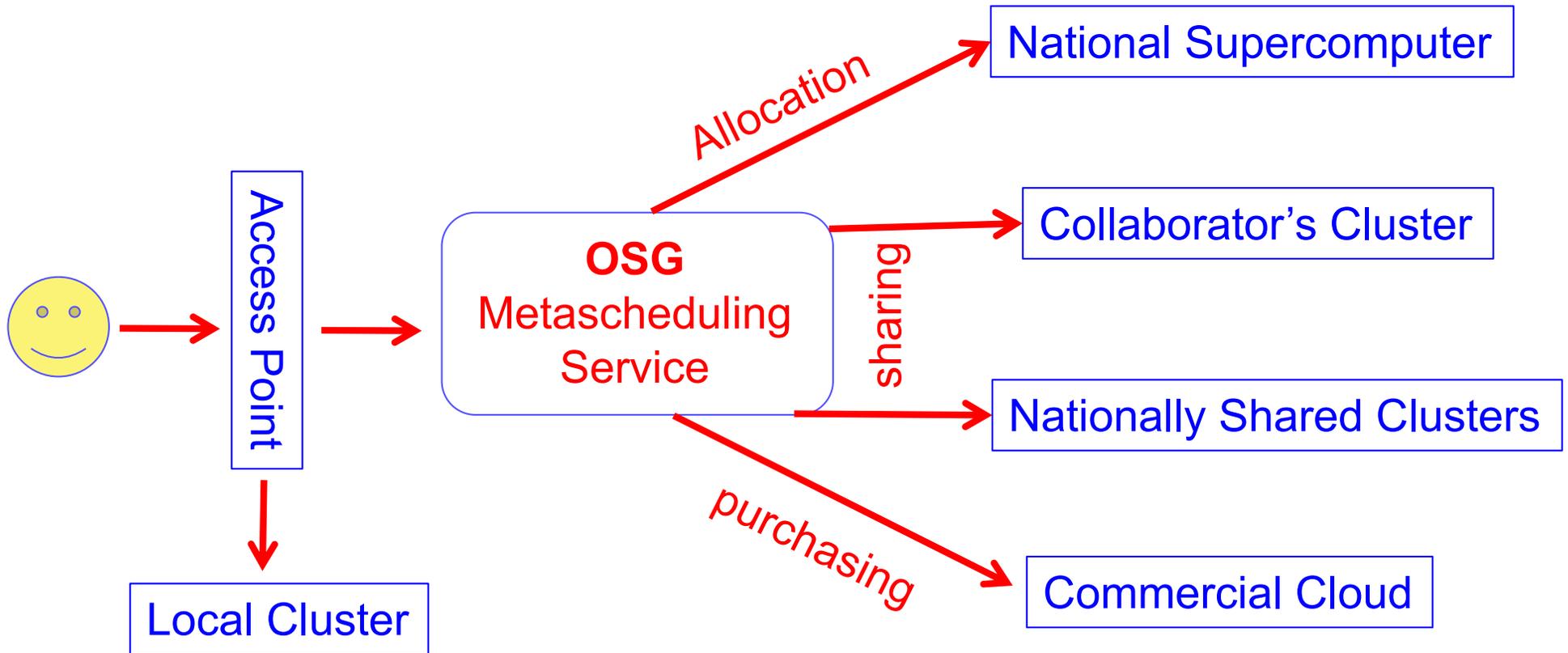
Resource Sharing via OSG

Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC

October 22nd 2020

Sharing compute resources
Sharing Data





CC* clusters may join this federation to share resources.



Open Science Grid

OSG Compute Federation



126 “green dots” listed on this map

Accounting for ~ 160M hours/month ... $160M / (24 \times 30) = 222$ thousand cores

... but the real story is so much more complicated

- OSG is dedicated to advancing science, and will do anything we can to integrate resources
 - **OSG-CE** deployed as interface to your cluster.
 - You decide who uses how much and when. **~ 1.8B/year**
 - **HTCondor “glideins”** join our resource pools.
 - Submitted to clusters without a CE, Cloud, **~70M/year**
 - **EGI-CE** submitted to via OSG operated gWMS **~0.8B/year**
- **Two types of resource pools**
 - Bring Your Own Resources = BYOR **~2.2B/year**
 - Open Pool of OSG **~400M/year**

- OSG works on three simple principles:
 1. **Resource Owners determine policy of use**
 - This means that all policy of use is set locally by the clusters that join the federation.
 2. **Resource Consumers specify the types of resources they are willing to use.**
 - How much RAM? How many cores per node? ...
 3. OSG submits its *own* batch system as 'jobs' into local batch systems.
 - **User jobs are submitted locally, queued centrally, and execute anywhere that matches requirements after resource becomes available.**

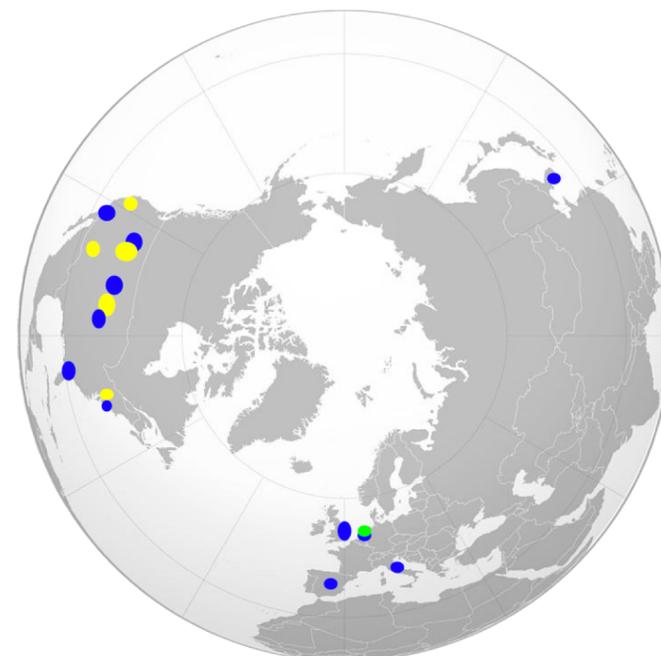
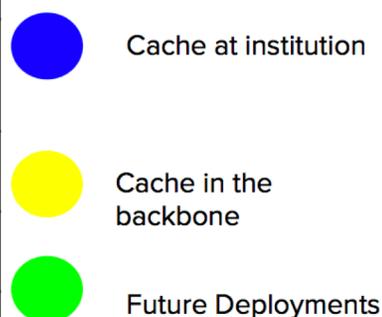
OSG operates overlay system(s) as services for all of science

OSG Data Federation

“Netflix for Open Science”

Collaboration	Working Set	Data Read	Reread Multiplier
DUNE	25GB	131TB	5.4k
LIGO (private)	41.4TB	3.8PB	95
LIGO (public)	4.3TB	1.5PB	318
MINERVA	351GB	116TB	340
DES	268GB	17TB	66
NOVA	268GB	308TB	1.2k
RPI_Brown	67GB	541TB	8.3k

**Data pulled from federation
in 6 month period 3-8/2020**



**More than a dozen caches
deployed across 3 continents**

Including several at POPs of Internet2 and GPN.

Recent use includes “collaborations”, “public data” accessed by single PIs, and “personal data” accessed by single PIs
(9 single PIs with >1TB of personal data read)



Data Federation Goals



- People come with their data on their local storage systems.
- OSG offers to operate a Data Origin Service to export your data into the OSG Data Federation.
 - We give you a globally unique prefix for your filesystem namespace, and then export your namespace behind it.
 - We allow you to decide who can access what.
- OSG then strives to guarantee **"uniform" performance across the nation by operating caches** to:
 - Hide access latencies
 - Reduce unnecessary network traffic from data reuse (by many jobs)
 - **Protect the data origins from overloads**

OSG operates overlay system(s) as services to all of science

Sharing does not imply “quid pro quo”

We are delighted to work with your campus to facilitate open science, in any form. Your researchers can utilize the OSG at any scale, regardless of how much computing you share with the community.

- OSG's objective is to “Advance Open Science through distributed High Throughput Computing”
- OSG thinks of its science stakeholders in terms of 4 categories:
 - Individual Researchers
 - Campus Research Computing Organizations
 - Multi-campus Science Teams
 - “Big Science” Collaborations
- OSG offers a diversified portfolio of services to support these different science stakeholders.

Contact us at: support@opensciencegrid.org