



State of OSG

Frank Würthwein OSG Executive Director UCSD/SDSC

June 2nd 2025







20th Anniversary of OSG





2005

4











Growth in 20 years

Institutions ~ 10x Data movement ~ 1Mx # of jobs/day ~ 1,000x





OSG as a Distributed Infrastructure

How do Institutions join today vs 20 years ago?



Different Service Options



- Compute
 - OSPool
 - More than a dozen Private Pools
 - From few 1,000 to few 100,000 cores per pool
 - "Pools are filled" in many different ways ...
 - gWMS submits to
 - Hosted CE
 - RPM installed CE
 - glidein container managed by site
 - User submitted glidein
 - Kubernetes provisioner
 - ... probably forgot some ...
- Data
 - Data Origins
 - Data Caches

Many Provisioning
Mechanisms to choose from.

Can join at different layers of the stack

- Join a Pool via a provisioning mechanism
- Join the NRP Kubernetes infrastructure
- Have PNRP operate your hardware from IPMI up



- Institutions differ in the effort they have to join, and the control they desire after joining.
 - Minimal effort & control => IPMI
 - Maximal effort & control => join pool(s) via hosted CE & gWMS







- Data Center
- Networking Infrastructure
- Hardware maintenance
- Cybersecurity
- System support
 - OS, batch system, storage system, middleware, …
- User support
- Management & budget

An example business model may look as follows:

Campus budgets 4FTE

Charge faculty ~40% of hardware costs for 6 years of operations.

 \Rightarrow ~\$2M/year in hardware to break even.

e.g. 100 nodes with 8GPUs each \$100k per node avg purchasing costs Nodes have 5 year warranty and are operated for 6 years, \$6667 per node/year in operations. ⇒100 x \$6667 = \$667k ⇒100/6 = 17 nodes to purchase/year ⇒\$1.7M hardware budget annually









207 Institutions integrate resources



International Institutions





Following the NSF rules, any US scientist may request that their international collaborators are integrated into OSG

For historic reasons, we do not count institutions that we engage with only because of the LHC





Open Science Data Federation



39 Institutions Contribute to OSDF Today





22 Origins and 37 caches across 5 continents



Roughly 40Gbit/sec averaged over the year



Fun Facts on Usage





- 26 top level directories with >1PB data read.
 - 17 OSPool users & LIGO (4), IceCube, NRP(2), KOTO, CHTC
- 40 top level directories with >1TB unique data read.
 - 6 OSPool users, 2 NRP users, CHTC & lots of collaborations: LIGO (9), LHCb, NRAO (4), EHT, XENON, JLAB, GlueX, REDTOP, Einstein Telescope, KOTO, SBND, SBN, Microboone, DES

Collaborations tend to have larger volumes of unique data. OSPool user access smaller volumes more often.



Open Science Store ... a new concept



- The CC* storage awards have a 20% contribution to the community requirement.
- We now have 4 such awards that have given us storage space to manage.
- We have started giving out allocations on that storage that meet NSF strategic goals
 - SAGE data
 - Internet Routing data
 - Internet Telescope data
 - Burn3D data commons
 - NOAA Fisheries data
 - Pelican Facilitation
 - NDP Facilitation
 - Astronomy data from IUCA
- This is strictly in prototyping stage ... we are not yet open for business more broadly ... contact us if you are interested in being a prototype user of OSStore.



Dare we predict the future?



- Reach 1,000 colleges with resources integrated
- Complete the token transition and "capability based authentication" not just within PATh services but globally.
- Strong growth in OSDF to support an open data ecosystem of many projects that build on each other and creatively compete with each other.
- Integration of instruments, sensors & IoT
 - Including their digital twins
- More compute & data in the Classrooms
 - Various types of AI as a service composed into distributed systems deployed via automated workflows
- More innovations in cybersecurity
 - A continuum from open to private to regulated

We will see more change during the next 20 years than the last 20 years





• OSG is celebrating its 20th birthday this summer.

- Scale increased between 10x and 1Mx depending on metric.
- Much more diverse services & ways to integrate into OSG today than even 5 years ago.

• We exceeded 200 institutions offering resources via OSG

- We are now accounting NRP institutions correctly

OSDF usage continues to grow in breadth

- Lot's of science collaborations
- Lot's of individual users
- Lot's of use from many pools, including OSPool
- Open Science Store is available in prototype mode









 This work was partially supported by the NSF grants OAC-2112167, OAC-2030508, OAC-1841530, OAC-1836650, the CC* program, and in kind contributions by many institutions including ESnet, Internet2, and the Great Plains Network.





