



Towards A National Research Platform that federates all academic Cyberinfrastructure in the USA

Frank Wuerthwein

Director

San Diego Supercomputer Center





The Minds We Need

- **Connect every community college, every minority serving institution, and every college and university, including all urban, rural, and tribal institutions** to a world-class and secure R&E infrastructure, with particular attention to institutions that have been chronically underserved;
- **Engage and empower every student and researcher** everywhere with the opportunity to join collaborative environments of the future, because we cannot know where the next Edison, Carver, Curie, McClintock, Einstein, or Katherine Johnson will come from; and



Long Term Vision



- Create an Open National Cyberinfrastructure that allows the federation of CI at all ~4,000 accredited, degree granting higher education institutions, non-profit research institutions, and national laboratories.
 - Open Science
 - Open Data
 - Open Source
 - **Open Infrastructure**
 - ← Open Compute
 - ← Open Storage & CDN
 - ← Open devices/instruments/IoT, ...?

Openness for an Open Society



**What do we mean by
"Open Infrastructure" ?**



The Power of Sharing



- Any participating institution shall be able to (dynamically) share any fraction of its resources with any other.
 - Collaborating Researchers can pool resources
- Institutions can share resources for the common good of all.
 - To democratize access funders can stipulate nationwide sharing for some fraction of the resources they fund.



Federation = distributed control



- Resource **owners determine policy of use** for what they own.
 - Implies policy is set locally at the resource.
- Resource **consumers determine policy of what they are willing to use.**
 - Implies policy is set locally at the user access point.
- Federated **system matches consumers to owners respecting both sets of policies.**



How can this possibly scale?



Scalability Challenge



- Resource Owners want minimal threshold for participation
 - And yet, have cybersecurity constraints that need to be met.
- Resource consumers want minimal threshold for participation.
 - And yet, have application constraints that need to be met.

**Massive challenge to design Cyberinfrastructure,
software, processes, and human teams,
that mediates this gap,
and scales to 1,000's of institutions.**



What's the State of the Art?



3 Ways to build Open Infrastructure Federation(s)



- At the container orchestration layer, e.g. K8S as implemented in Pacific Research Platform.
- Federating independent container orchestration frameworks, e.g. via SLATE or via Admiralty on K8S.
- Federating storage and compute clusters at the storage and batch system layer., e.g. via OSG.

**More (cybersecurity) control
implies more effort to join & operate.**

Some Institutions will never be able to join because of a mismatch between the effort they have and the control they desire.



Complementarity in Implementation of “Bring Your Own Resource” model

OSG focused on campus cluster integration.

Pacific Research Platform focused on individual node integration instead of clusters.

In the following we introduce these models



Pacific Research Platform



- “Bring Your Own Resource” philosophy
 - From “soup kitchen” to “potluck” supercomputing
- Pioneered integration at the K8S & IPMI layer
 - Integrating resources at the single node level
 - Provide updated hardware shopping list for institutions
 - 2021: 11 different options from 8-way GPU, to 4-way FPGA, to 648TB SAS, ...
 - Among the most popular are 8-way GPU and 648TB Ceph storage nodes
 - join at K8S layer (institutions responsible for their own base OS install)
 - Or allow central team to devOps base OS via IPMI from UCSD/UNL
- Global hardware integration across 30++ institutions
 - 11 MSIs (8 in CA) and 6 institutions in EPSCoR states
- User interfaces: Jupyter, kubectl, OSG, HTCondor
- Community dominated by CS and EE
 - 500++ GPUs & many PBs of disk space for machine learning



The OSG Model



- Institutions operate their own compute & storage clusters.
- OSG provides software & services that allow integration of clusters
 - Overlay batch system to create compute federation
 - Data integration to create data federation
 - All data in data federation is accessible from compute in compute federation.



OSG Compute Federation



149 “organizations” listed on this map

Compute Resources at 64 institutions in the USA contributed last year.



Institutions contributing to OSCF



- **64 US institutions contributed compute resources** during the year ending on March 9th 2022.
 - We count institutions, organizations inside institutions & “clusters”
 - E.g. UCSD is an institution with 2 “green dots”: SDSC & CMS group in Physics. 6 clusters contributed resources in 2021, incl. one entirely inside the commercial cloud, and another that itself has hardware across 30 institutions.
- Out of these 64 institutions in the USA
 - **9 are Minority Serving Institutions (MSI)**
 - 1 CC*, 2 EPSCoR, 1 Non-R1
 - 10 are in EPSCoR states
 - 17 received an NSF CC* award for a compute cluster
 - 13 are non-R1
 - 31 are none of the above

26 of the 64 US institutions are either MSI, EPSCoR, or non-R1

OSG is Democratizing Access to Cyberinfrastructure



OSG Data Federation (OSDF)



20 Caches ... 6 of which are in R&E network backbone

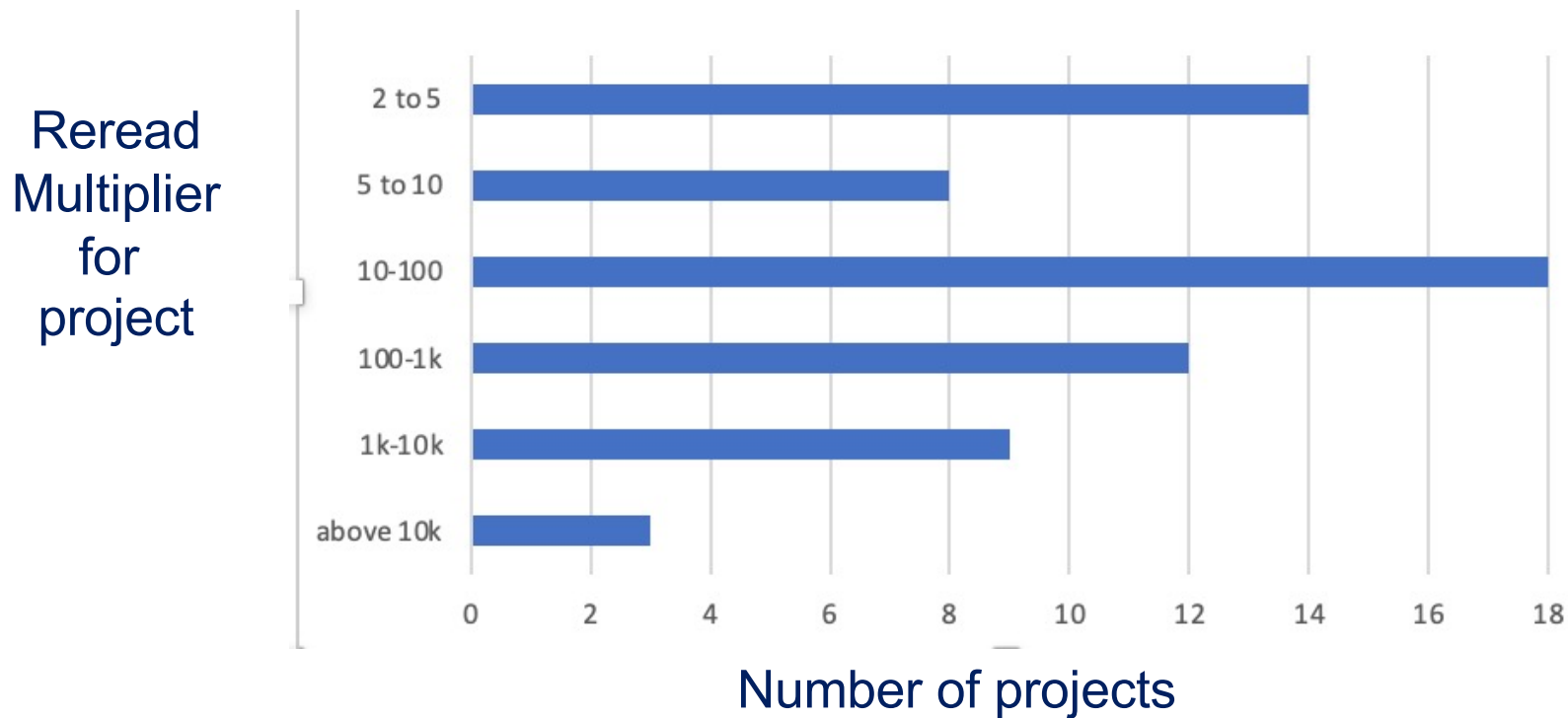
10 Data Origins ... incl. one in Europe



The OSG Data Federation



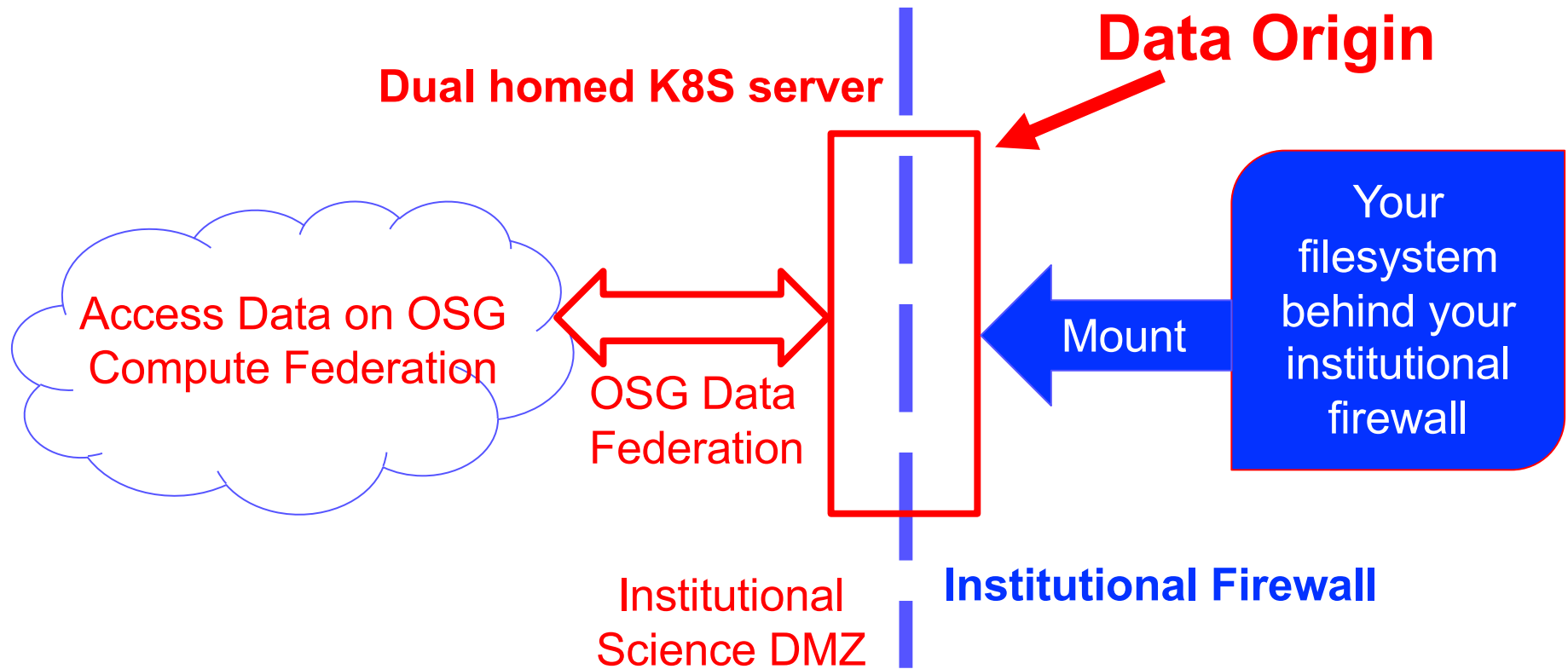
**In 2021,
92 research groups, 9 collaborations, 1 campus
read 32PB of data out of a working set of 420TB
for an average re-read factor of 75.**





You Can Join your data in 2 ways

- 1) Transfer your data via HTTPS to the PATH supported Origin
- 2) **Federate the filesystem at your institution with the OSDF**



Your fileserver with your data can be behind your institutional firewall. A dual homed K8S server mounts only those filesystems you want to export. We operate the OSDF API by deploying a “Data Origin” container into K8S.



Functionality of a Data Origin



- **Export your data read-only into the Data Federation**
 - You choose what part of your filesystems namespace you want to export.
 - You can change this dynamically any time you want.
 - **Data can be public or private**
 - Origin uses HTTPS as protocol => works as general webserver in addition to OSG Data Federation.
- **Store output data produced on OSG**
 - Put via HTTPS as part of HTCondor workflows
 - authorized only to those people you want to support.
 - Read-only access possible to data stored this way.
 - What is put into an origin may be read via the federation if desired.



OSDF's Global Namespace



- Global Namespace is separate from the origins that hold the data
 - **You can move data between origins via HTTPS without changing how the data is accessed via the OSDF.**
 - Literally, nobody will notice !!!
- This allows federation of namespace that is separate from federation of server hardware that serves the namespace.
 - Lot's of interesting ways of using the power this provides you with.



Top Users of Data Federation



Project	Data Read	Working Set	Reread multiplier
LIGO (Private)	10PB	38TB	264
Minerva	5.6PB	3.1TB	1,789
NOVA	2.6PB	1.9TB	1,348
LIGO (Public)	2.4PB	38TB	67
Tufts_Hempstead	2.0PB	380GB	5,321
DUNE	1.6PB	185GB	8,658
Steward	1.0PB	11TB	92
REDTOP	874TB	95TB	9.2
Molcryst	570TB	5GB	115,650
BiomedInfo	530TB	66GB	8,090

17 projects have >TB working sets

11 of these are individual researchers and small groups

Tufts Computer Architecture Lab

Steward Observatory Data Analytics

R&D towards future particle physics experiment

Quantum chemical and machine learning insights into supra-molecular organization of molecular crystals

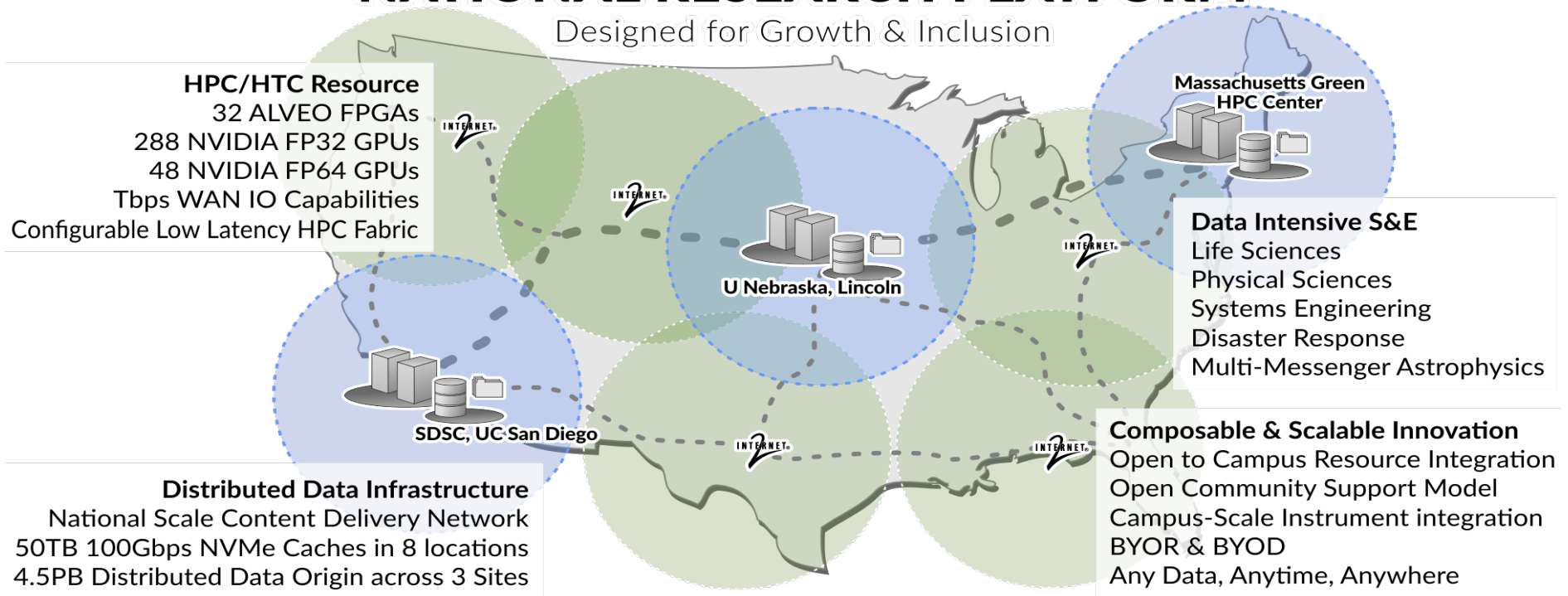
Development and application of software tools for performing large-scale biomedical informatics on microbial genome sequence data.



What's next?

NATIONAL RESEARCH PLATFORM

Designed for Growth & Inclusion



Nationally Distributed Category-II System (NSF)

3 year prototyping phase followed by 2 year allocation phase.



Data Infrastructure Model of NRP



- **Support regional Ceph storage systems across the USA.**
 - Campuses can join individual storage hosts to the Ceph system in their region.
 - All regional storage systems are Origins in OSG Data Federation (OSDF)
 - Deploy replication system such that researchers can decide what part of their namespace should be in which regional storage.
- **Deploy caches in Internet2 backbone such that no campus nationwide is more than 500 miles from a cache.**
- Between all the collaborating projects, we ought to have >10PB of data origin storage in OSDF to support open science by the end of 2022.

NRP is a much more experimental data infrastructure than OSDF.

Filesystem mounts of regional storage on compute in the region.

User controlled replication of partial namespaces across region.

Institutional BYOR into regional storage systems.

NRP caches are part of the OSDF.



Summary & Conclusion



- It takes a cyberinfrastructure ecosystem of collaborating projects to build and operate the open infrastructure that will achieve our joint vision of democratizing access for all.
- **We are encouraging all of you to join us!**
- To achieve our collective vision we can use all the help we can get ! Let's coordinate and collaborate.



Acknowledgements



- This work was partially supported by the NSF grants OAC-2112167, OAC-2030508, OAC-1841530, OAC-1836650, and MPS-1148698, CNS-1456638, CNS-1730158, ACI-150112, ACI-1541349, OAC-1450871 (NU), OAC-1659169 (SDSU)

