



State of OSG

Frank Würthwein
Director, San Diego Supercomputer Center

January 16th 2024





OSG “Statement of Purpose”

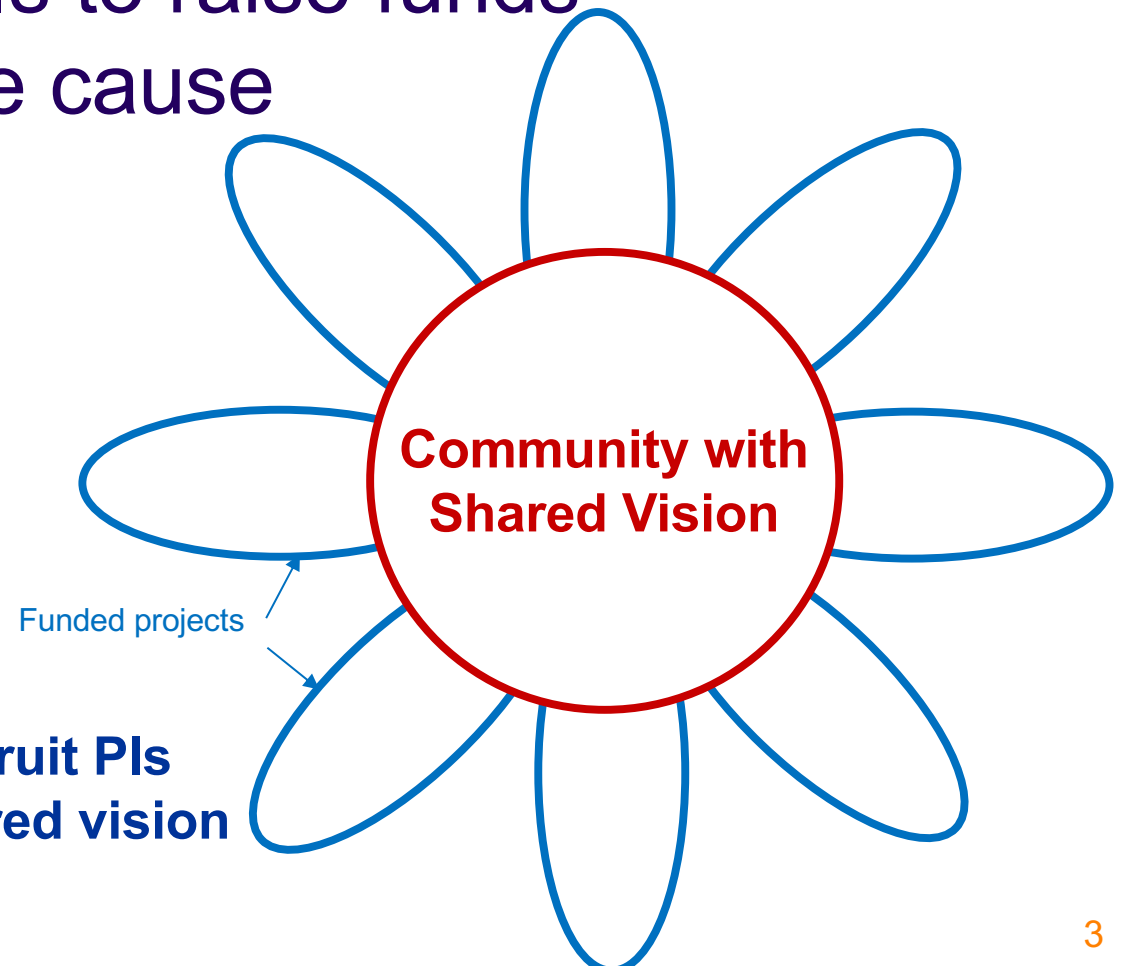
OSG is a consortium dedicated to the advancement of all of open science via the practice of distributed High Throughput Computing (dHTC), and the advancement of its state of the art.



What does this mean?



- OSG is not a project with funding
- OSG is open to PIs to raise funds to contribute to the cause



It's up to the council to recruit PIs to expand or change the shared vision



Some projects have been building the foundations



- Partnership for Throughput Computing (PATH)
 - Provides core funding ... some details to follow
- IRIS-HEP
 - Provides core funding ... see details by Derek Weitzel



Partnership for Throughput Computing (PATH)

PATH brings together ... the Center for High Throughput Computing (CHTC) and the Open Science Grid (OSG) Consortium. (From NSF awards notice)

I'm taking information straight from the [PATH website](#).



Bring the power of throughput computing to your science with **PATH**

If you are a **Researcher**

Sign up for OSPool Capacity

or

Apply for PATH Compute Credits

If you are an **Organization**

Join us at PATH for Organizations

Our Software

PATH is focused on researchers and organizations other than the LHC

For Researchers



New! The [PATH facility](#) provides compute capacity made available for researchers by the NSF through request. The dedicated capacity allows researchers to run HTC workloads on purpose built hardware from the convenience of a dedicated Access Point.

[Apply for Compute Credits](#)

[View Documentation](#)

US-affiliated researchers can use an OSG-Operated Access Point to harness the capacity of the [Open Science Pool](#). Sign up now and join the hundreds of [projects](#) using the OSPool!

[Sign up on the OSG Portal](#)

For Organizations

Build local HTC capacity with the [HTCondor Software Suite](#). Share computing capacity, data, and software between collaborators and even across continents.

Share campus computing and data resources into the [US national cyberinfrastructure](#) via OSG. Contact our support staff to collaborate on the right options for your desired outcomes.

[Contact Support Staff](#)



Services



Credit Accounts

The Partnership to Advance Throughput Computing (PATH) project is funded by the NSF to advance High Throughput Computing (HTC) and its impact on research and education. To achieve this goal, PATH operates a fabric of capacity services that ranges from Research Computing Facilitation to Access Points capable of managing distributed HTC workloads. PATH has added to its fabric of services support for capacity credit accounts.

PATH's Support for Research Computing and the CC* Program

Final Deadline Passed: September 11th, 2023

Production Services

The Production Services team, as part of the Fabric of Capacity Services (FoCaS), maintains services essential to delivering science using OSG Consortium resources. The Compute Management services allow for sharing of computing resources, provisioning pools of computing resources, job execution management. Data Management services provide for serving and pre-placing data in support of workflows. Additionally, services such as identity management, monitoring, and accounting enhance and secure the fabric.

Facilitation Services

Facilitation Services leverage the CHTC-pioneered principles of Research Computing Facilitation to accelerate dHTC uptake by campus researchers and collaborations via the Open Science Federation and OSG-Operated Access Points, and by campuses and other organizations interested in advancing their own dHTC and dHTC Facilitation capabilities.

Software Development

The Software Development Team supports, sustains, and enhances the HTCondor Software Suite (HTCSS) to enable the potential of distributed High Throughput Computing. This suite of software tools includes HTCondor, components to build both on-premise HTCondor compute clusters for use by independent academic, commercial, and government campuses, and components to support the federation of processing and data resources across the OSG.

Global Infrastructure Laboratory

PATH is committed to openness in the science and institutions we support, in innovation of technologies and methodologies, and in interfaces to the broader ecosystem of NSF-funded CI services, especially as the fabric of coordination services evolves. To facilitate ingestion of ideas into HTCSS, and services into FoCaS, the Global Infrastructure Lab (GIL) tests and evaluate infrastructure software from inside and outside the partnership.



Science Collaboration Support



1. REDTOP
2. IceCube
3. IGWN
4. LSST-Duke
5. South Pole Telescope
6. VERITAS
7. XENON
8. SNOWMASS21
9. KOTO
10. CLAS12
11. GLUEX
12. EIC
13. MOLLER
14. UHE Neutrino Observatory Trinity
15. EUSO-SPB2
16. DUNE
17. EHT
18. NRAO

Mission:

Help midsize science collaborations leverage the OSG fabric of services.

In addition to services provided to individual researchers and campuses via PATH, we also provide Rucio and FTS.



State of PATH in maps



HTCSS Users





OSG Deployment Map



Counting of institutions is incomplete because of NRP



OSG Storage vs Compute





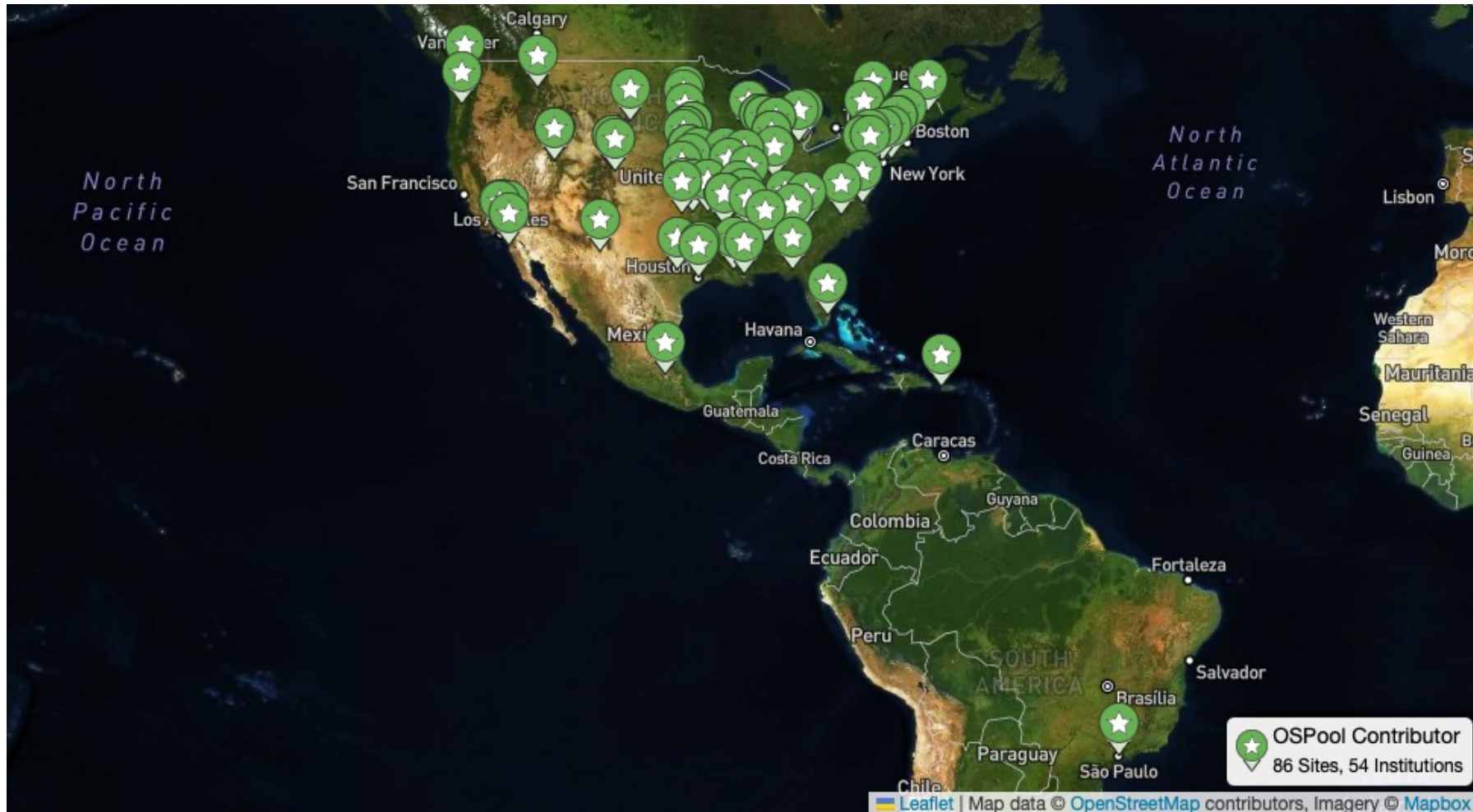
Open Science Data Federation



42 sites across 32 institutions



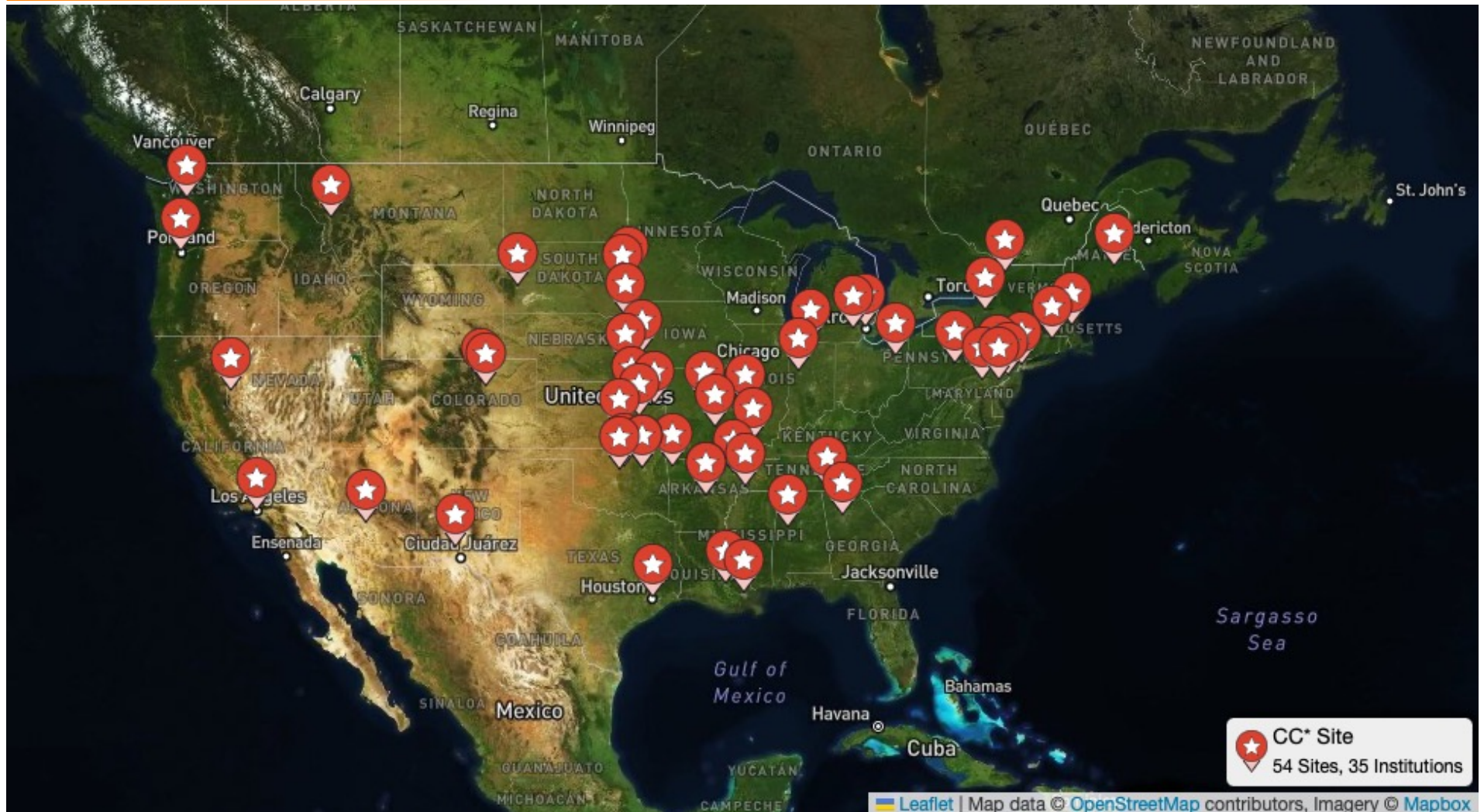
OSP Pool



Counting of institutions is incomplete because of NRP



CC* Awardees



Counting of institutions is incomplete because of NRP



Pelican: New Kid on the Block



Pelican: Advancing the Open Science Data Federation Platform

Brian Bockelman
Principal Investigator
Morgridge Institute for Research



Miron Livny
Co-Principal Investigator
University of Wisconsin-Madison

Frank Wuerthwein
Co-Principal Investigator
University of California San Diego

This project is supported by the  under Grant OAC-2331480.



The Pelican Project

The OSDF is operated by  using hardware from  and others.

Who develops the software?

The Pelican project (OAC-2331480) has the following goals:

1. Strengthen and Advance the OSDF.
2. Expand the types of computing where OSDF is impactful.
3. Expand the science user communities.
 - With a particular driver of the climate community.



Year 1 Activities

Development Activities:

- Re-engineering the origin service to drastically simplify deploys.
- Simplifying the OSDF onboarding / registration process.
- Deploying new storage usable by the NDC-C community.
- Developing new clients for Python- and browser-based user.
- Partnering with the National Data Platform Pilot to provide indexing/metadata services.

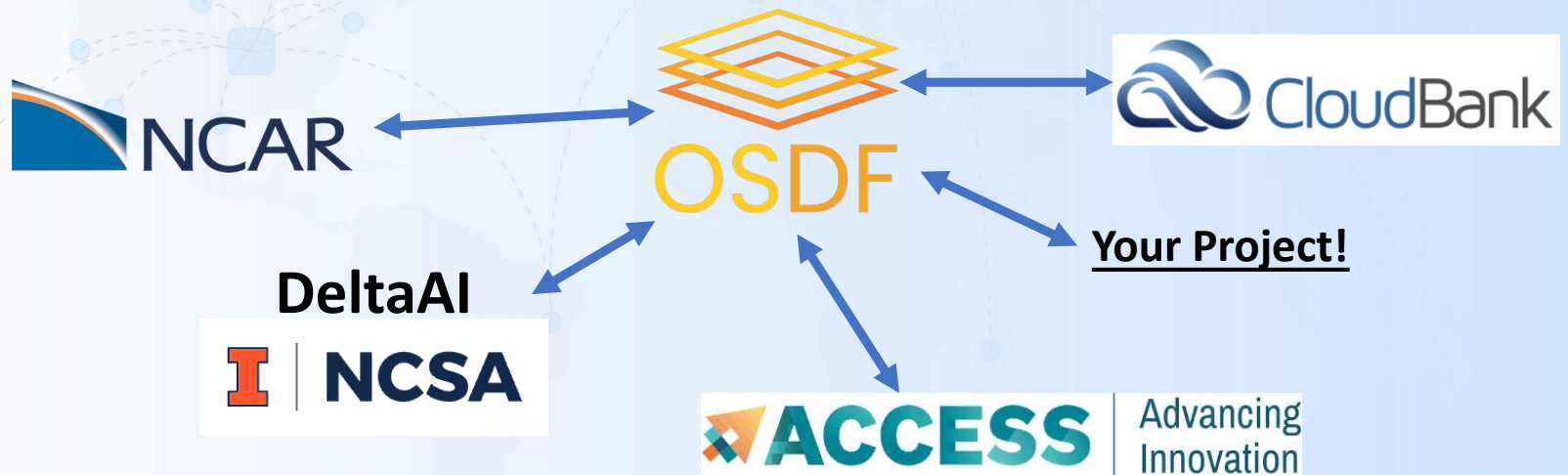
Integration & Community Activities:

- Connecting OSDF into computing resources represented in NDC-C.
- Integrating climate-related AWS Open Data datasets into OSDF.
- Developing OSDF training that we can give at your workshops.
- Hosting student fellows over the summer.



OSDF and High Throughput Computing

Most of our users utilize OSDF to serve High Throughput Computing workloads. Through NDC-C, we aim to expand the OSDF impact.



NDC-C is a great venue to collaborate and amplify the impact of these existing resources

NDC-C = National Discovery Cloud for Climate



OSDF by the numbers

Over the last 12 months, the OSDF transferred

**64_{PB} &
62 files/s**

Data used by

**15 science collaborations &
~120 OSPool users**

How can we achieve 10,000 users of OSDF?

I want to count datasets accessible and in use.



Vision



- Have other projects build on top of PATH & Pelican.
 - National Data Platform builds a federated data catalogue following FAIR principles for all of science on top of OSDF.
 - Fusion Data Platform for AI builds on top of OSDF.
 - National research Platform builds underneath and OSPool and OSDF, and will in the future build on top.



Aside on NRP Architecture



- Depending on effort available and control desired, you can build on NRP both vertically and horizontally at different layers of the stack.



- NRP is a non-local extendable container deployment platform, thus allowing many uses unthinkable for a SLURM cluster in a data center.



Summary & Conclusion



- Growth is the #1 priority
 - Growth in under-resourced institutions joining “the maps”.
 - Growth in projects in OSPool, including researchers from under-resourced institutions.
 - Growth in (diversity of) data accessible via OSDF.
 - Growth in projects that build on our infrastructure.