



# Expanding Services for Campuses

**Lauren Michael**  
**Research Facilitation Lead**  
**UW-Madison**

**Frank Würthwein**  
**OSG Executive Director**  
**UCSD/SDSC**

**March 15, 2022**





# Brief Overview of Concepts



Researcher connects to  
compute & data resources  
via an access point

Access Point

Institutions connect  
compute resources via a  
**Compute Entrypoint (CE)**

Compute resources are  
aggregated as **Pools** in the  
Open Science Compute  
Federation (OSCF)

Jobs running on a resource  
pool access data via  
**Data Caches**  
in the Open Science Data  
Federation (OSDF)

Institutions connect  
data resources into the OSDF  
via a **Data Origin**

**We will walk you through these services in this talk.**



# Compute Resource Pool



- In general, an organization need not operate their own pool, unless:
  - You are a network of campuses that want to share resources across the network.
  - You are looking for advice on how to deploy your own *local* cluster (e.g. via HTCondor).
- In that case, talk with us, and we provide advice appropriate to your specific circumstances.



# Compute Entrypoint



- **Contribute your local cluster/cloud into an existing pool** (e.g. the Open Science Pool)
- **Requirements:**
  - Outgoing network from cluster nodes.
  - (OSG-hosted CE) Provide OSG with an account on your cluster, accessible via SSH keys.
  - Maximize use and reduce network impact by installing Singularity and software for data caching.
- **We take care of the rest. Details at URL below.**

<https://opensciencegrid.org/docs/compute-element/hosted-ce/>



# Functions of an Access Point



- provide accounts to people, and define their privileges
- manage job execution across configured resources:
  - pool(s) in the OSCF, local cluster, national resource, and/or commercial cloud, etc.
  - a local HTCondor cluster's access points can be configured to integrate the Open Science Pool, for example
- integrate access to locally-staged data
- manage data access across resources
  - export local data into the OSDF
  - expose other data in the OSDF that the user has access to

**Many institutions send researchers to OSG Connect, with OSG-managed access points.**

**If you're interested in a 'local' AP, talk to us, and we will provide advice tailored to your specific circumstances.**

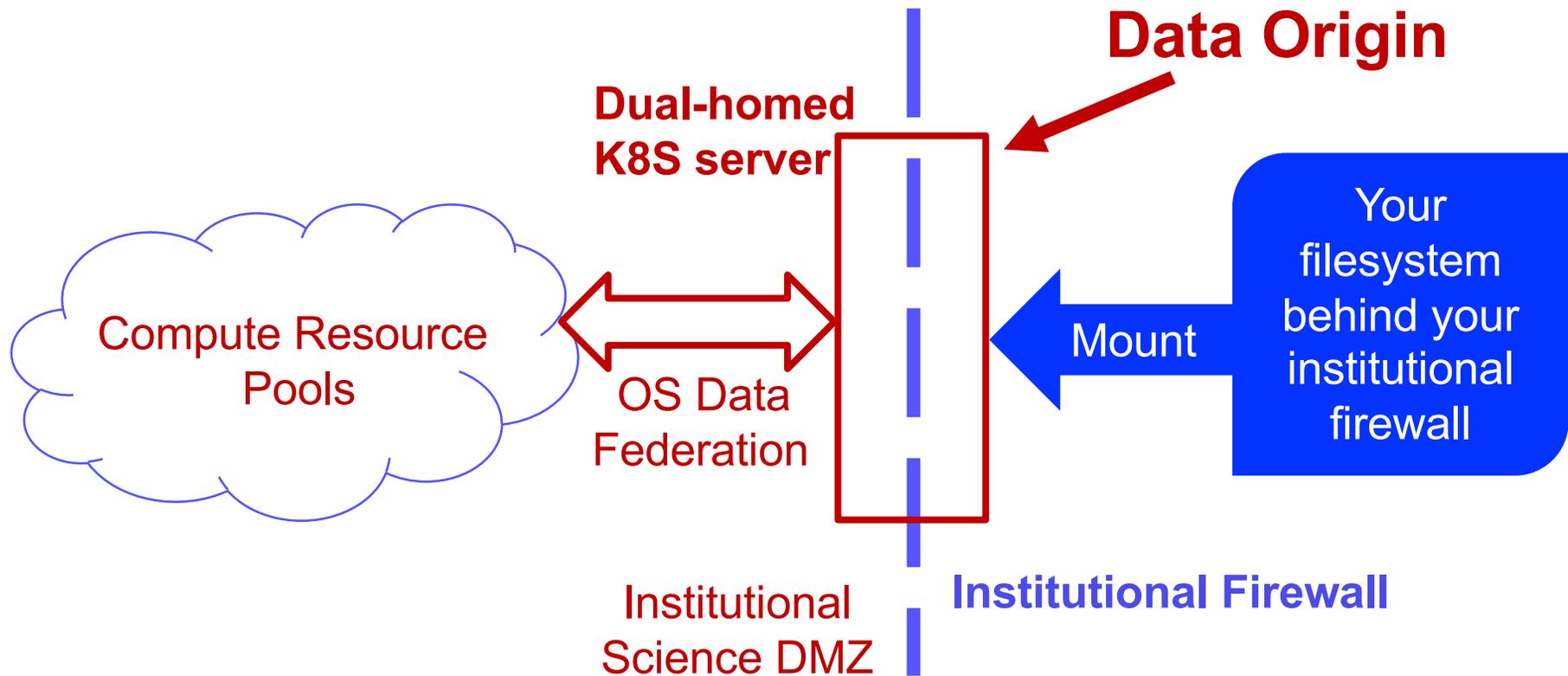


# OSG Connect



- OSG-operated service for US-based projects
  - <https://www.osgconnect.net/> > 'Sign Up'
- Includes
  - Access Points integrated with the **Open Science Pool**, and with the option to integrate other user-specific resources (e.g. allocations, cloud)
  - **Data storage** integrated with the OSDF, via the OSG Connect Data Origin
  - OSG Research Computing Facilitators provide **personalized onboarding**, email-based **support**, virtual **Office Hours** (today!)
  - **Documentation, software** solutions, and regular **User Training**:
    - <https://support.opensciencegrid.org/support/home>

**We love to co-Facilitate alongside campus staff!**



Your fileserver with your data can be behind your institutional firewall. A dual-homed K8S server (operated by OSG) mounts only the filesystems you want to export.



# Functionality of a Data Origin



- **Export your data read-only into the Data Federation**
  - You choose what parts of your filesystem's namespace are exported, and make changes as you wish.
  - **Data can be public or private.**
  - Origin integrates with OS Data Federation and as a general-purpose web server (via HTTPS).
- **Store output data produced on OSG**
  - You control 'write' access authorization.
  - Put via HTTPS as part of HTCondor workflows
  - Data written to a Data Origin can be automatically exported via the OSDF (as above), if you desire.



# Data Cache



- OSG and partners operate regional Data Caches across the OSDF.
- If we haven't already deployed a cache in your region, we are likely to do so in 2022.
  - Most batch systems that have full outgoing network connectivity generally do not need to deploy a cache.
- If you are concerned that you have a large site, with limited outgoing network connectivity, talk with us to discuss options.
  - Cache deployment and operations can be very similar to the data origin (e.g. k8s containerization, operated by OSG).

- **OSG User School**

- Week-long in-person program resuming summer 2022
  - Great for researchers and staff who support their computing needs
  - **Applications opening soon. Watch for emails!**
- Past Materials: <https://opensciencegrid.org/outreach>

- **Presentations and Trainings**

*targeted for campus IT/research computing staff, gateway developers, etc.*

- **dHTC Campus Workshops**

- <https://indico.fnal.gov/event/45998/>
- <https://indico.fnal.gov/event/46925/>

- **Site Admin Office Hours (today!)**

- External presentations and trainings
  - Past: RMACC 2021, PEARC21, Gateways 2020, etc.





# Summary



- For most campuses, an **OSG-hosted CE** and **Data Origin** are sufficient for full integration into OSG.
- **OSG Connect** is available to any US-based project.
- Local **Access Points** support local user identity and seamless integration of local data.
- New compute resource **Pools** may be relevant to networks of campuses wanting to share resources.

Talk to us about any of the above via [support@opensciencegrid.org](mailto:support@opensciencegrid.org) and check out various learning opportunities.



# Acknowledgements



- This work was partially supported by the NSF grants OAC-2030508, OAC-1841530, OAC-1836650, and MPS-1148698

