Building Communities Around OSDF and OSPool Contributors

Christina Koch · OSG Researcher Facilitation Lead
Tim Cartwright · OSG Campus Coordinator

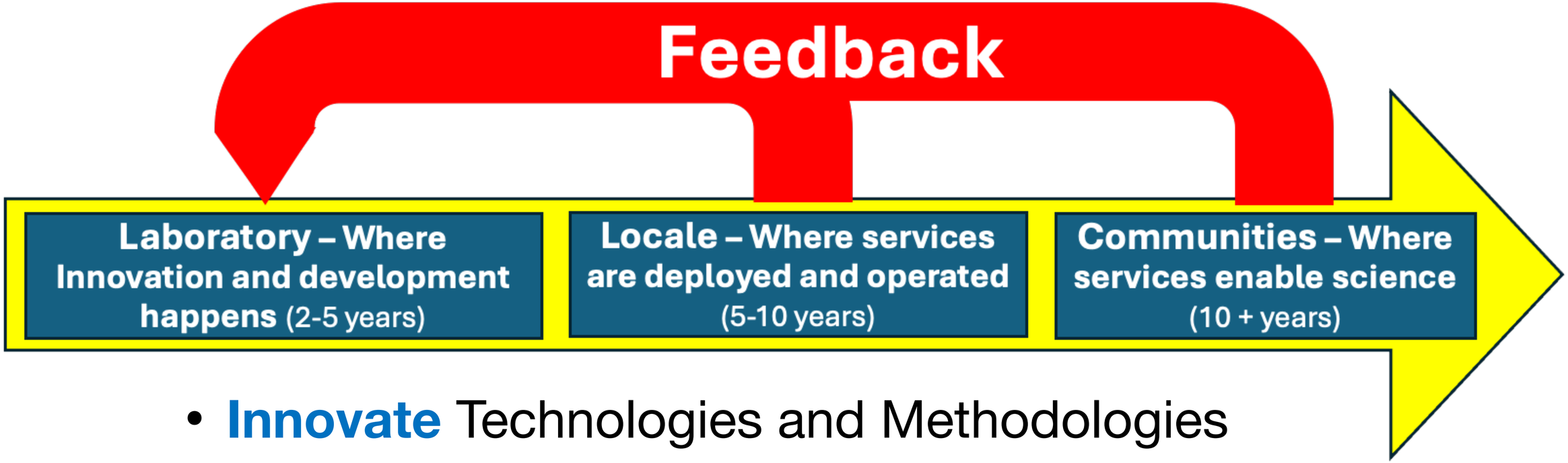
Center for High Throughput Computing and PATh Project
University of Wisconsin–Madison

HTC25 – Community
4 Jun 2025

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

OSG

Reminder: Translational (Computer) Science



- **Innovate** Technologies and Methodologies
- **Deploy** Services
- **Build** Communities

If You Remember One Thing…

Building the community of contributors is an inseparable aspect of providing the OSDF, OSPool, and OSStore.

Without the lab and locale, the community have nothing to use.

Without a community, the lab and locale have no direction
(“if you build it, they will come”).

Once the community exists, the lab and locale must continue supporting it.

The rest of this talk consists of various examples of this in practice

Direct Community Building

Talking to Campuses

- Key **community building** activity: having conversations with campuses
- Ways to talk to us
  - Planning and integration meetings
  - Weekly campus meet-up
  - Events like HTC25, presence at meetings/conferences
- How conversation builds **community**
  - Develops trust relationships
  - Effective way to work through technical processes
  - Conversation is also a tool for discovering **feedback**

Other Community Connections

- Connecting campuses to the community can also happen asynchronously
  - Contact form on the website [https://path-cc.io/contact/]
  - Support email [support@osg-htc.org]
  - Slack workspace [opensciencegrid.slack.com]
  - Website and documentation [https://osg-htc.org/campus-docs/]

New Community: Storage

Where We Have Been

- **OSPool Contributions**
  - Many moving pieces to contribute successfully to the OSPool.
  - Circa 2019, campuses started arriving with CC* computing grants - now what?
  - Contribution process centralized with Tim as campus coordinator.
  - The process has evolved based on interactions with the **community**.
- In order to integrate new campuses to the contributor **community**, we had to go from bespoke and brittle contribution processes to a well-defined and repeatable contribution process.

Where We Are Going

- **Open Science Data Federation**
  - The OSDF was a "one-off" data delivery platform,
    - Tightly coupled to the OSPool.
    - Limited tool and infrastructure development.
  - Pelican: aims to formalize the data federation tooling
  - Circa 2024, campuses arriving with CC* Storage awards - now what?
- **Let's learn from previous experience!!**
  - Establish a point person (currently Christina).
  - Build a storage contribution process based on the compute contribution process.
- Moving forward, we consider storage and compute contributors as part of the same community.

Sidebar: Defining the OSStore

- As we started working with campuses contributing storage, we needed to refine our language.
- Campus are not contributing to the OSDF per se; they are contributing **storage** that is made accessible via OSDF services.
- Some campuses are contributing storage that we (PATh) can delegate to groups for sharing their data sets.
- That collection of contributed storage now has a name: **OSStore**

Integrating Storage into OSDF/OSStore

- **Initial meeting: goals and implementation details**
  - **What does "contribution" look like?**
    - Storage behind a Pelican cache service
    - Storage that has existing datasets, shared via a Pelican origin service
    - Storage that we (PATh) can delegate to groups who want to share datasets (OSStore)
  - **How will the needed Pelican services be operated?**
    - Operated by the campus
    - Operated by NRP
  - **How much storage is available?**

New Capability: Hosting Data in OSStore

- What to do with the OSStore? Host data!
- Current process (work in progress):
  - Get in touch via email (support@osg-htc.org)
  - Initial meeting to discuss:
    - Goals of data sharing
    - Data size and use patterns
    - Level of access
  - Implement OSDF services and get the data into the OSStore.
- See what we have so far here: https://osg-htc.org/services/osdf/data

Feedback Drives OSPool Changes

Refining OSPool Integrations

- Old news (~2½ yrs): Developed a streamlined process for OSPool integrations
  - Kickoff meeting → preparations → live integration meeting
  - Added focus on working together (which builds community)

- **Feedback:** Campus staff want written list of requirements
  - Helpful to prepare for new integration
  - Handy at other times, too; *e.g.*, after significant site changes
- We developed and posted a checklist:
  - https://osg-htc.org/campus-docs/ospool-requirements/
  - Currently for Hosted CE & Slurm — will add other cases (esp. container)

New Capability: CE Dashboard

- **Feedback:** Campus staff and leadership want information on contributions and researchers/institutions being supported
- HTCSS team developed the ***CE Dashboard***
  - Demoed at HTC24, in production now for (PATh-)Hosted CEs
  - https://ce-dashboard.ospool.osg-htc.org/
- Todd T. (main dev) will demo in Campus Track this morning
- Next steps:
  - Add campus-hosted HTCondor-CEs
  - **Get feedback:** How might you use this? What could be improved?

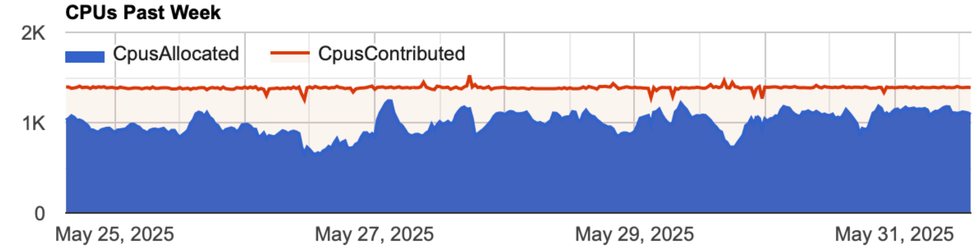
HTCondor View

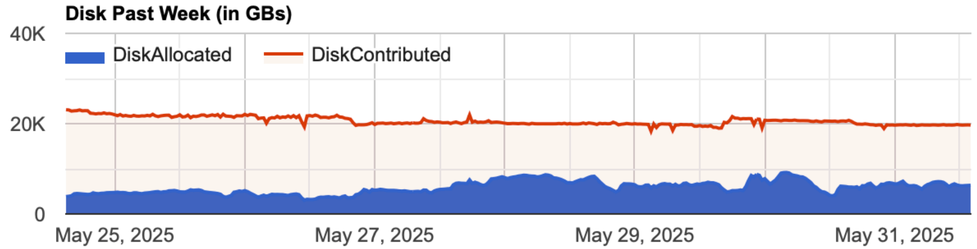
University of Wisconsin–Madison

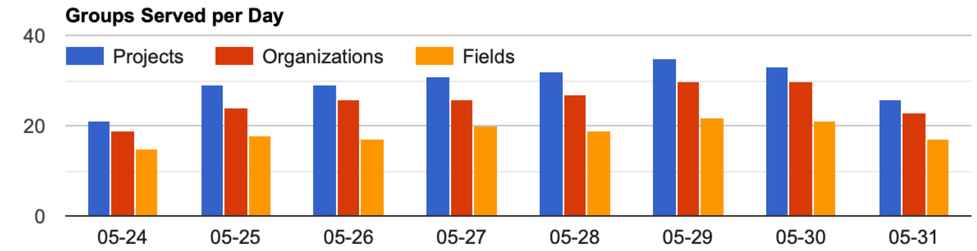
Overview Contributed

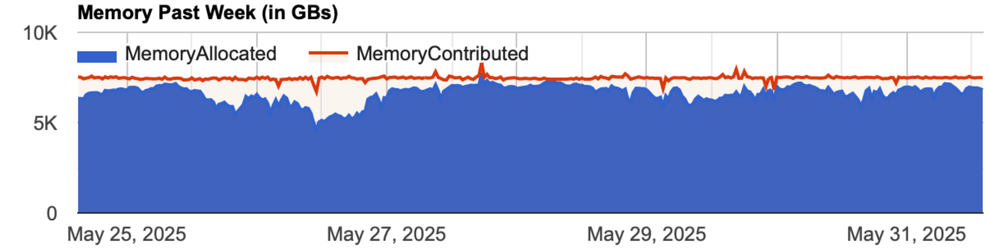
Contribution of compute capacity from the University of Wisconsin–Madison cluster named CHTC.
Description of this Compute Entrypoint (CE): Hosted CE serving CHTC Spark Cluster

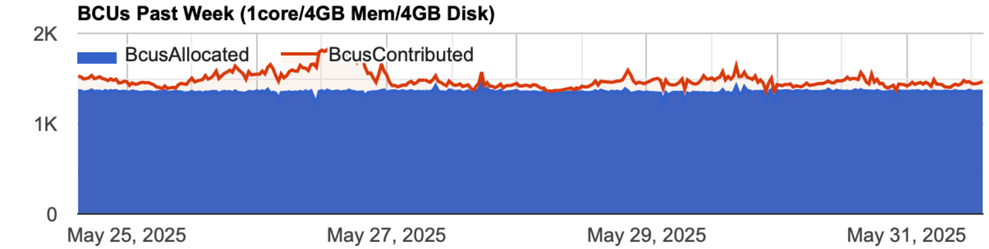
Last Update Time	CE Health	Projects	Fields	Organizations	Cpus
2025-05-31 14:00	Good	14	11	13	516

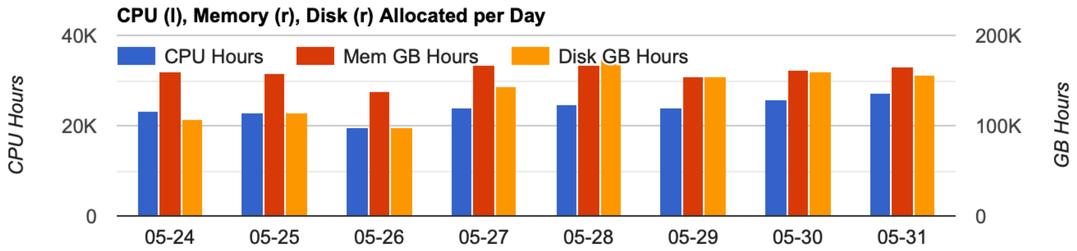












Proactive Monitoring for the Community

OSPool monitoring

- Identify, triage, and take action on issues in under 1 week from onset
  - Did something change with pattern of contributions? Why?
  - Are contributions being used? How? Are things changing?
- CE Dashboard has improved internal methodology for detecting issues

OSDF monitoring

- Pelican team does a lot — they do Translational CS, too, and OSDF is primary **Locale**
- I have started doing some, esp. watching for effects on the OSPool

Please give **feedback**: There is always room for improvement!

Feedback ⇒ 4 New OSPool Features

Singularity/Apptainer Detection & Validation

- **Feedback:** Does my site need to provide Apptainer?
  (because 75–90% of researcher jobs run in containers)
- GlideinWMS developers: No – now a glidein can use HTCondor's copy:

1. Site-specific overrides that PATH staff add to GlideinWMS config
2. Normal locations (where "normal" is defined by GlideinWMS)
3. The HTCondor install within this glidein
4. module load singularitypro *or* singularity
5. Other uncommon locations

Glidein Scratch Directory Cleanup

- **Feedback:** Some glidein scratch directories are left behind, filling disk!
- PATh team developed a clean-up process
  - Occurs during glidein start-up process; safely cleans as much as it can
  - Deployed in February
  - Since then, maybe 1–2 reports of filled disks due to glideins
- **New feedback:** Need something similar for OSPool EP containers
  - In testing now
  - Plan to deploy in next week or two

OSPool EP Container Image Updates

Feedback: Want to contribute ***GPUs*** using the OSPool EP container

- Matt W. (Software team) added this

```
docker run -it [...] \
  -v /etc/OpenCL/vendors:/etc/OpenCL/vendors:ro \
  hub.opensciencegrid.org/osg-htc/ospool-ep:24-cuda_11_8_0-release
```

Feedback: Does the EP Container really need `--privileged`?

- Matt W. (working with Greg T.) are finding ways to reduce needed privileges

```
docker run -it[...] \
  --security-opt seccomp=unconfined --security-opt systempaths=unconfined \
  --security-opt no-new-privileges \ [...]
```

https://osg-htc.org/docs/resource-sharing/os-backfill-containers/

The condor_who Tool in Glideins

Feedback: At times, site admins want to know about OSPool researcher jobs

- TJ (HTCSS developer) updated condor_who tool to work with glideins
- Sample output (edited to fit, etc.):

PROJECT	USER	AP_HOSTNAME	JOBID	RUNTIME	MEMORY	DISK	CPUs	EFCY	PID	STARTER
KOTO	****	ap23.uc.osg-htc.org	27872550.7518	0+00:47:33	2.0 GB	6.0 GB	1	0.99	3000808	2999526
MIT_submit	****	submit06.mit.edu	764122.27	0+00:02:01	1024.0 MB	2.3 MB	1	0.99	3169488	3168950
PixleyLab	****	ap20.uc.osg-htc.org	35396737.1095	0+06:04:38	4.0 GB	4.0 GB	1	0.99	2052820	2049582
SSGAforCSP	****	ap21.uc.osg-htc.org	17594309.420	0+16:48:15	14.8 GB	4.0 GB	1	0.99	2802678	3838451
PSU_Lynch	****	ap20.uc.osg-htc.org	35421713.0	0+05:24:08	20.0 GB	2.0 GB	1	0.99	2302554	2301237
BiomedInfo	****	ap20.uc.osg-htc.org	32684753.98178	0+00:22:17	15.6 GB	32.0 GB	1	0.99	3066645	3066091
PixleyLab	****	ap20.uc.osg-htc.org	35396944.129	0+05:52:25	4.0 GB	4.0 GB	1	0.99	2199648	2165608

https://osg-htc.org/campus-docs/#monitoring-and-information

Summary

Be Part of the Community!

Building the community of contributors is an inseparable aspect of providing the OSDF, OSPool, and OSStore.

Reach out to us:

- On the OSG Consortium Slack workspace – ask for invite, if needed
- Join the OSDF-OSPool Campus Contributors mailing list – ask us how
- Join the Campus Meet-Ups – Wed., 11am Central (alt. weeks in summer)
- Email support@osg-htc.org

For more, join the HTC25 Campus Track next (upstairs, Room 201)

Questions?

This work is supported by NSF under Cooperative Agreement OAC-2030508 as part of the PATh Project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.