Open Science Grid

Submit locally, run globally.

# Supporting Your Researchers to Use OSG

Lauren Michael, Facilitation Services Lead

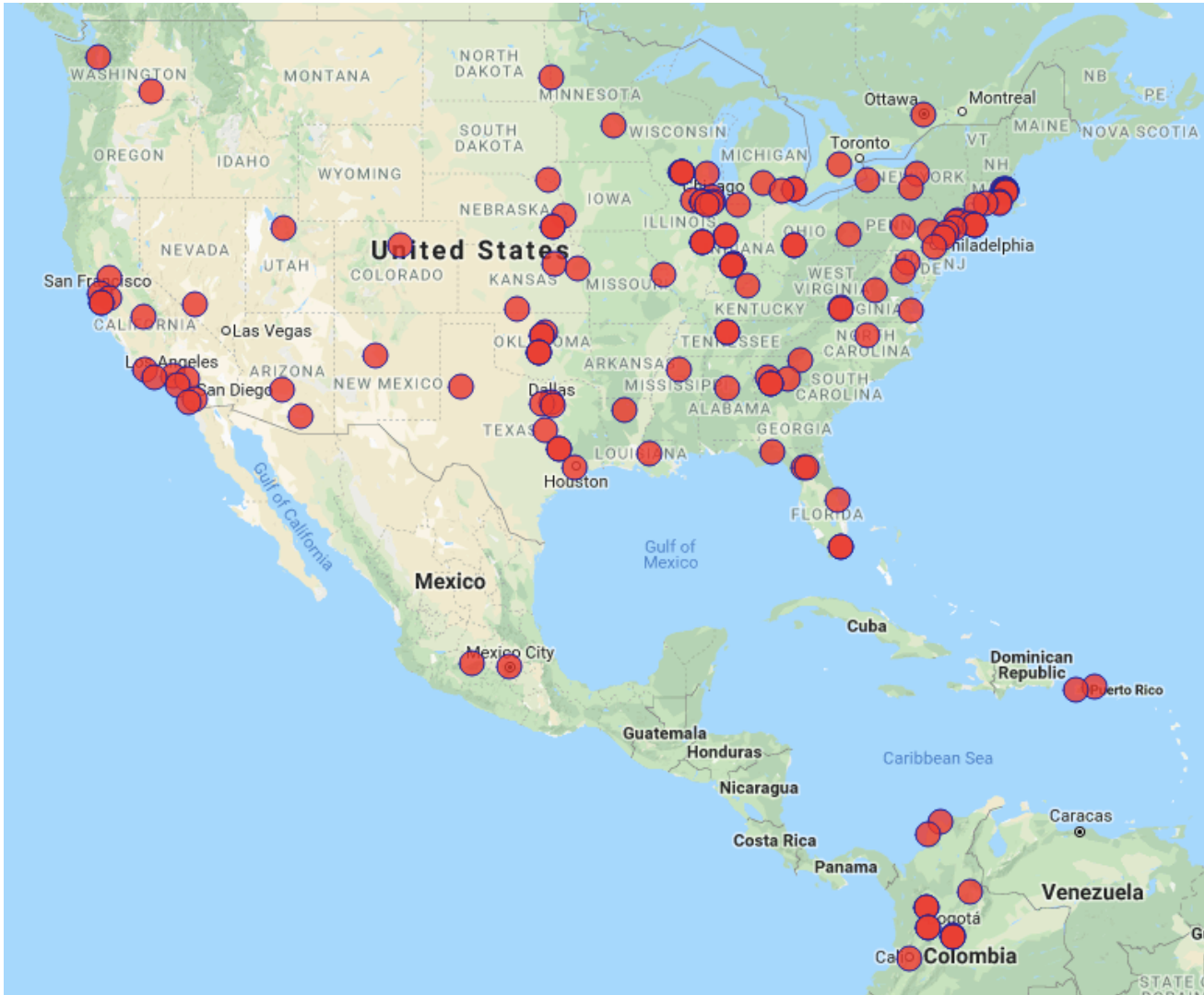
Open Science Grid

- Access to OSG via OSG Connect
- Learning Opportunities for Researchers and Campus Staff
- ‘Local’ Submit Points into OSG

# Who Participates?

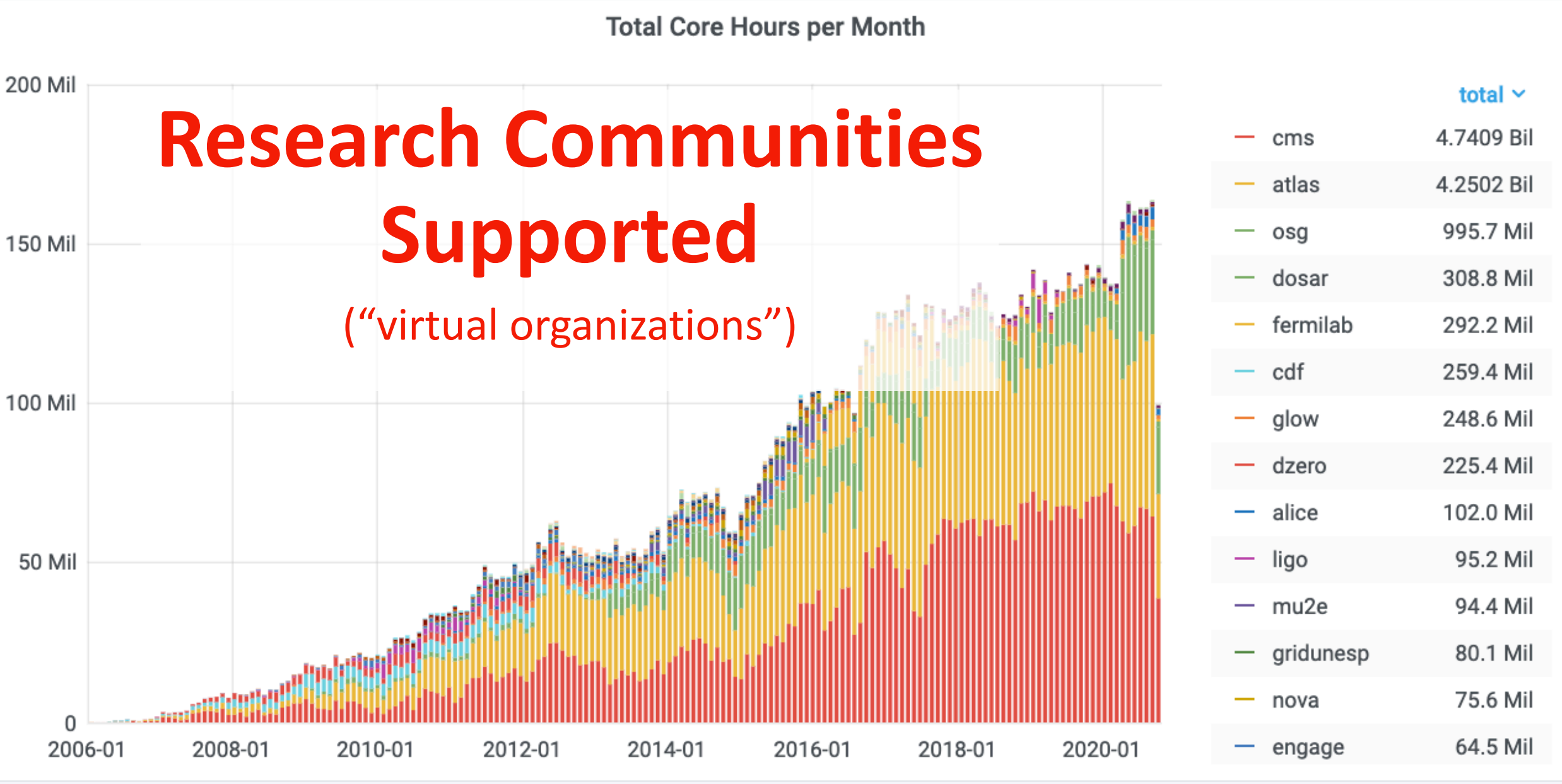
- Individual researchers and groups
- Multi-Institution Collaborations
  - Atlas/CMS (Higg Boson), and others
- Academic Institutions and National Laboratories
- Science Gateways

*Advancing this spectrum requires a diversified portfolio of services.*

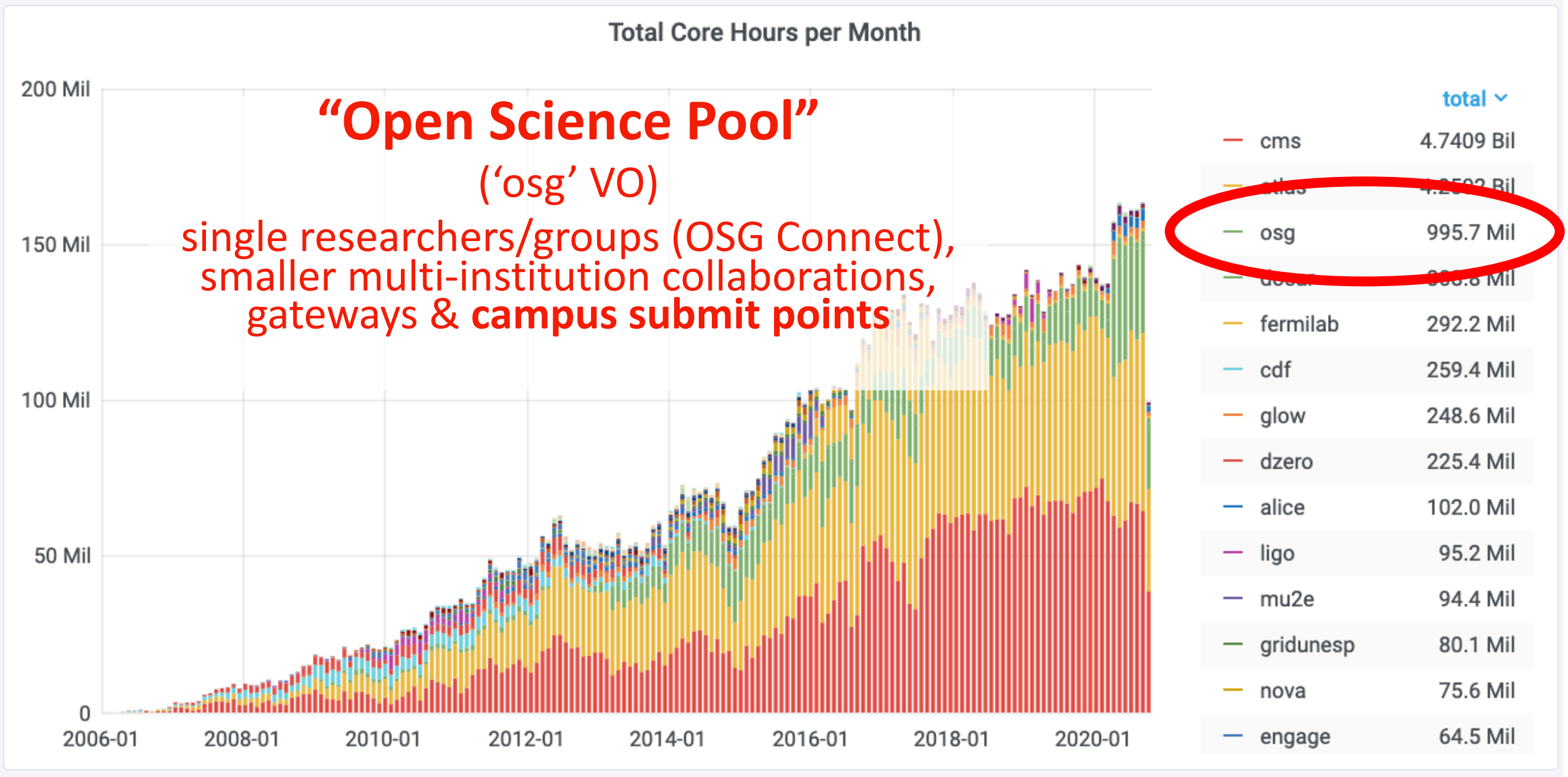


# Research Communities Supported

("virtual organizations")




https://gracc.opensciencegrid.org




**“Open Science Pool”**

(‘osg’ VO)

single researchers/groups (OSG Connect),
smaller multi-institution collaborations,
gateways & **campus submit points**

https://gracc.opensciencegrid.org

Open Science Grid

# Open Science Pool

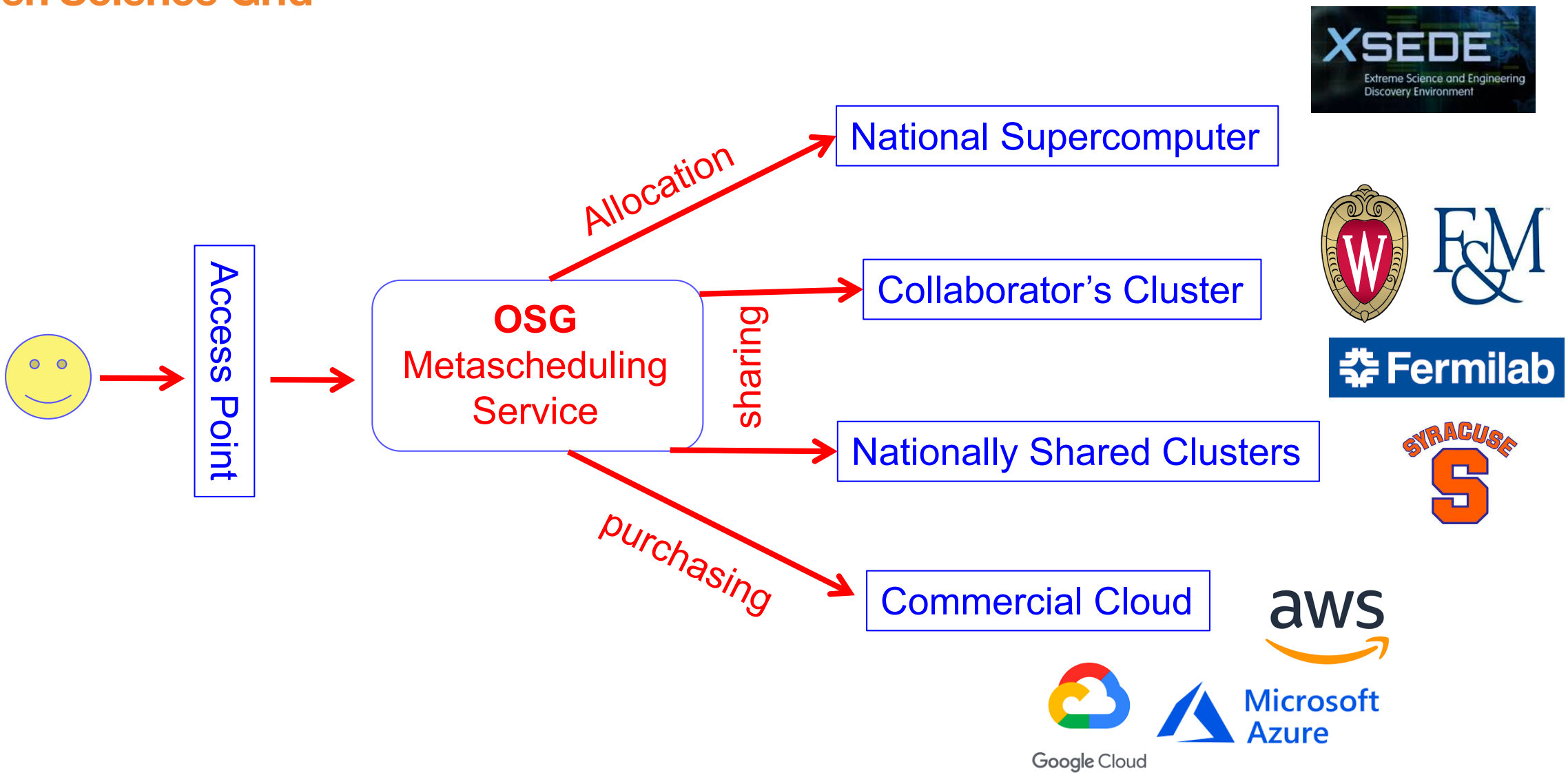
## Core Hours by Field of Science

| | total |
|---|---|
| Biological Sciences | 109.7 Mil |
| Physics | 78.5 Mil |
| Astronomy | 29.1 Mil |
| Chemistry | 22.7 Mil |
| Engineering | 5.22 Mil |
| Integrative Activities | 3.35 Mil |
| Agricultural Sciences | 2.787 Mil |
| Health | 2.187 Mil |
| Other | 1.888 Mil |
| Mathematics | 1.223 Mil |
| Computer Sciences | 958 K |
| Education | 500 K |
| Economics | 87.4 K |
| Earth and Ocean Sciences | 15.3 K |

## Core Hours by Institution

| | total |
|---|---|
| Folding@Home Consortium (FAHC) | 68.9 Mil |
| Massachusetts Institute of Technology | 30.7 Mil |
| Stanford University | 21.2 Mil |
| University of Pittsburgh | 19.6 Mil |
| Rochester Institute of Technology | 16.7 Mil |
| University of Hawaii at Manoa | 14.61 Mil |
| Lancaster University | 9.37 Mil |
| University of Chicago | 8.13 Mil |
| Fermilab | 6.45 Mil |
| University of Wisconsin-Madison | 6.06 Mil |
| University of North Carolina at Chapel Hill | 5.90 Mil |
| Wayne State University | 5.87 Mil |
| University of Pennsylvania | 5.81 Mil |
| Arizona State University | 5.23 Mil |
| University of Arizona | 4.86 Mil |
| New Mexico State University | 4.36 Mil |
| Georgia Institute of Technology | 3.65 Mil |
| Duke University | 3.17 Mil |
| LSU School of Public Health | 2.986 Mil |
| Wichita State University | 1.795 Mil |

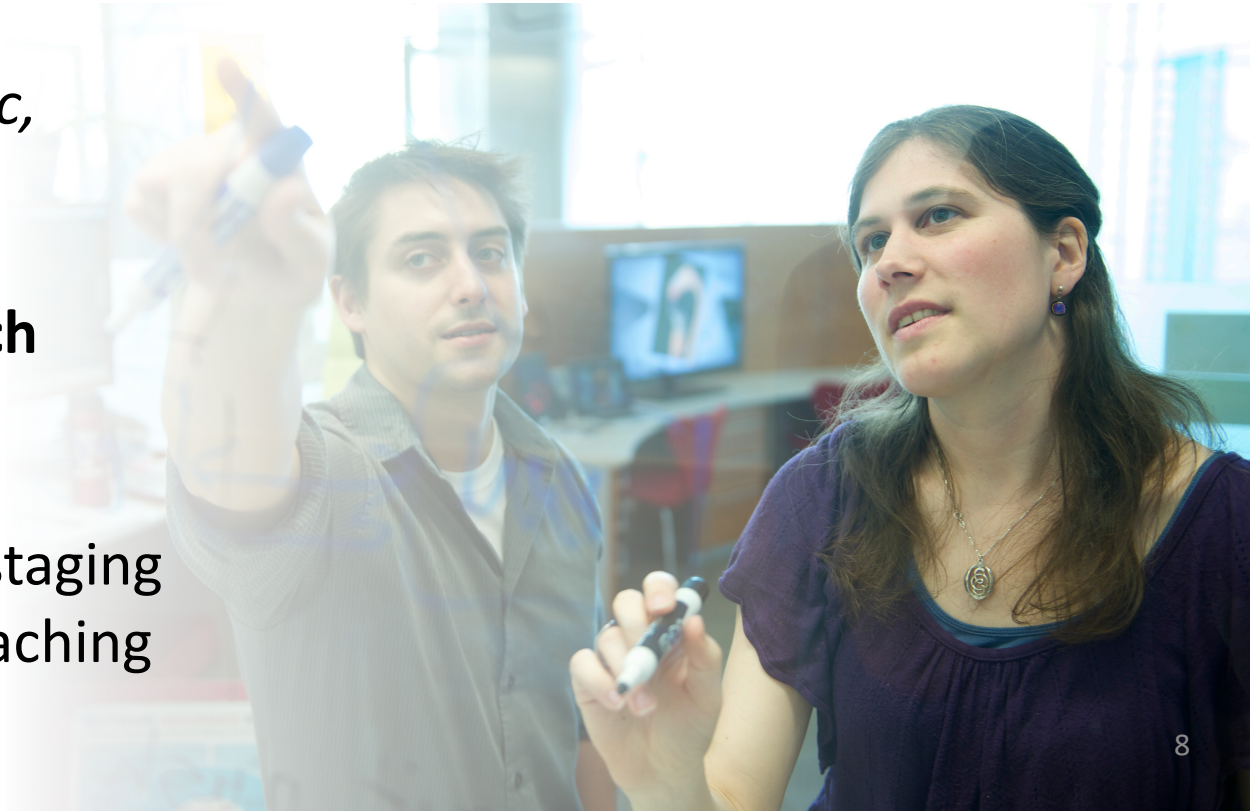
# Services Overview



# OSG for researchers: **OSG Connect**

***(FREE) access to OSG and dHTC Facilitation, with OSG-supported access points***

- **osgconnect.net > "Sign Up"**
- *available to researchers at any U.S. academic, government, or non-profit organization*
- includes:
  - **initial consultation with an OSG Research Computing Facilitator**
  - online documentation and examples
  - (roughly) unlimited scratch; support for staging software and large input w/ automatic caching
  - no "allocation" or proposal required



Open Science Grid

# For individual researchers: **OSG Connect**

OSG CONNECT

Support Login Sign Up

OSG CONNECT

Front door to the Open Science Grid

Documentation here

Free computing power for US researchers

Login Sign Up

# OSG Connect (in the last year)

Open Science Grid

### Total Wall Hours

**209 Mil**

### Total Wall Hours for Top 10 Projects

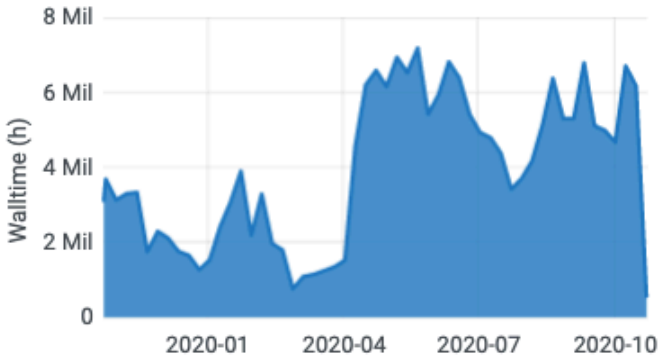
| | total |
|---|---|
| COVID19_FoldingAtHome | 68.9 Mil |
| COVID19_Stanford_Das | 21.2 Mil |
| CompBinFormMod | 16.7 Mil |
| Hawaii_Doetinchem | 14.61 Mil |
| chemml | 10.24 Mil |
| LancasterPPS | 9.37 Mil |
| REDTOP | 6.41 Mil |
| WSU_3DHydro | 5.77 Mil |
| molcryst | 5.25 Mil |

### Total Jobs

**61.2 Mil**

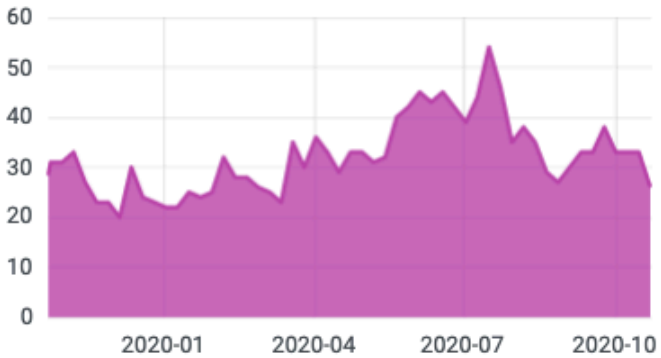
## Time Series

### Wall Hours per 7d



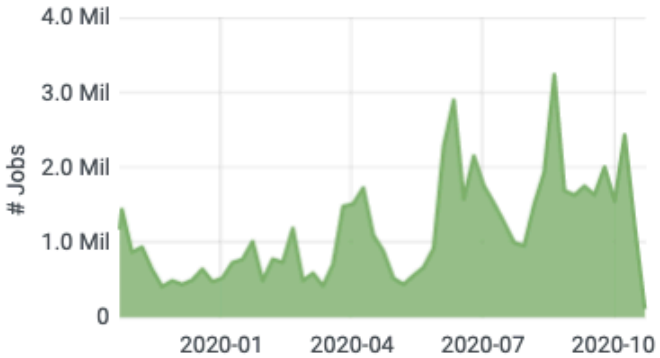

Sum CoreHours Total: 208.58 Mil

### Active Projects per 7d




Unique Count ProjectName Avg: 32

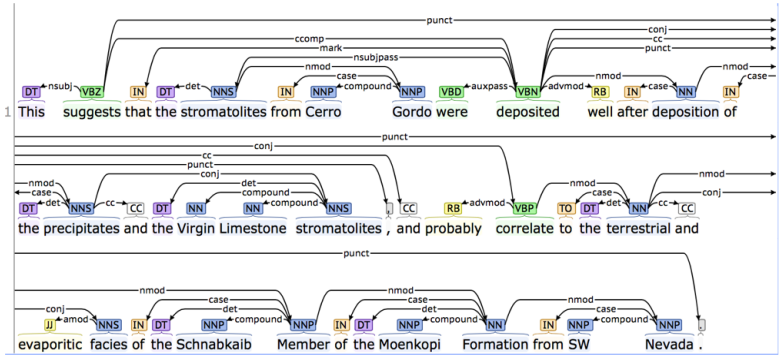
### Total Jobs per 7d



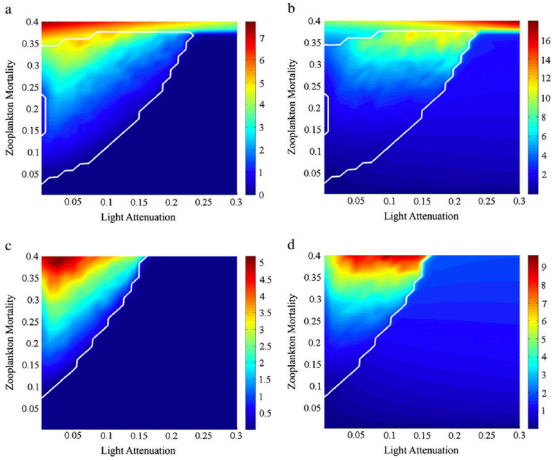

Sum Count Avg: 1.134 Mil

# Is it HTC-able?

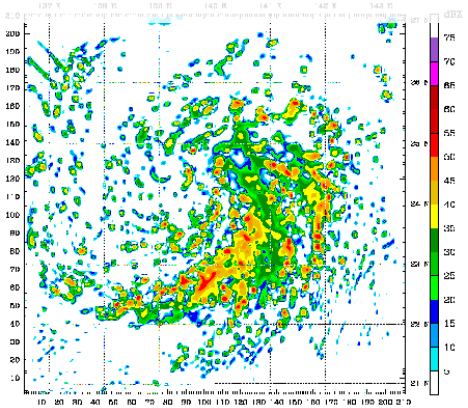
Is your problem potentially computable as lots of independent pieces? For example …



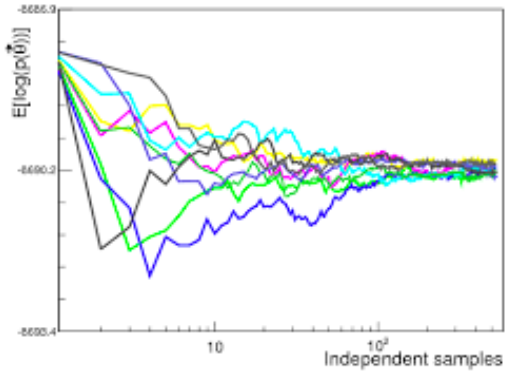

text analysis (most genomics …)



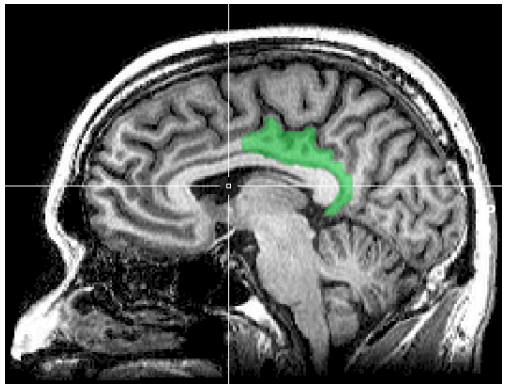

parameter sweeps



multi-start simulations




statistical model optimization (MCMC, numerical methods, etc.)



(multi-)image and sample analysis

# Signs of HTC-able work

- Any mention of **numerous** samples, images, models, parameters, etc.
- Nearly anything **written by the researcher** (e.g. c/fortran, Python, R)
  - Break out of loops!
  - Common internal parallelism could really be HTC (e.g. Matlab's 'parfor', 'distributed server', etc.)
- Some community softwares that use **multi-threading or multiprocessing** (e.g. OpenMP)
  - many are simply looping over data portions or independent tasks
  - HTC-able: break up input (or 'parameter' space), turn off multi-threading, combine results
- **Long-running** jobs; see above explanations

# Is it OSG-able?

Is your work computable as lots of *laptop-sized* pieces?

Open Science Grid

| ***Per-Job*** **Resources** | **Ideal Jobs!** (up to 10,000 cores, per user!) | **Still Very Advantageous!** | **Probably not…** |
|---|---|---|---|
| **cores** (GPUs) | **1** (1; non-specific) | **<8** (1; specific GPU type) | **>8 (or MPI)** (multiple) |
| **Walltime** *(per job)* | **<10 hrs*** *or checkpointable | **<20 hrs*** *or checkpointable | **>20 hrs** |
| **RAM** *(per job)* | **<few GB** | **<10 GB** | **>10 GB** |
| **Input** *(per job)* | **<500 MB** | **<10 GB** | **>10 GB** |
| **Output** *(per job)* | **<1 GB** | **<10 GB** | **>10 GB** |
| *Software* | *'portable' (pre-compiled binaries, transferable, containerizable, etc.)* | *most other than →→→* | *licensed software; non-Linux* |

# Research Computing *Facilitation*

***accelerating research transformations***

proactive engagement

personalized guidance

teach-to-fish training

technology agnosticism

upward advocacy

collaboration liaising

# Research Computing Facilitators

## The Missing Human Link in Needs-Based Research Cyberinfrastructure

ECAR

ECAR Research Bulletin | May 16, 2016

Lauren Michael, University of Wisconsin–Madison
Bruce Maas, University of Wisconsin–Madison

## Overview

As the roles of core, campus-supported IT services for research have expanded—including the emergence of cloud-based models—the benefits of on-campus human support and user engagement have become increasingly apparent. Ongoing challenges in securing research funding reemphasize a need to demonstrate significant societal impact via effective and efficient investments. At the same time, many campus research computing providers still face challenges in engaging researchers represented in the "long tail" of computing needs, where potentially significant, compute-enabled transformations to scholarship have yet to be realized. The most common models for research computing resources may already meet the significant needs of well-established or "traditional" users, typically in the physical

# What's Different about OSG Facilitation?

Open Science Grid

1. **HTC/dHTC is frequently new to users**
   - 'splitting up' work, optimizing throughput (over per-job performance), etc.
   - many have HTC-able work and don't know (you may not, either)
2. **OSG job logistics are different than using local resources**
   - built-in data movement instead of shared filesystem
   - software portability tools instead of system-wide installation
   - inherent interruption/retry
   - testing and troubleshooting on non-local resources
3. **Potential scalability and research transformation!**

# HTC Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**

# HTC Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**

- *Tell me about your research …*

  *… and how does computing fit in?*
- *What is your near-term bottleneck?*
- *How are you running it now?*

  *(the compute/data requirements?)*

  *(your computing background?)*
- ***How much/big would you like to run?***

# HTC Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
3. **Set expectations.**

If you …

> execute your work ***this way*** (HTC)

> requiring ***these learning steps*** and ***this much work***.

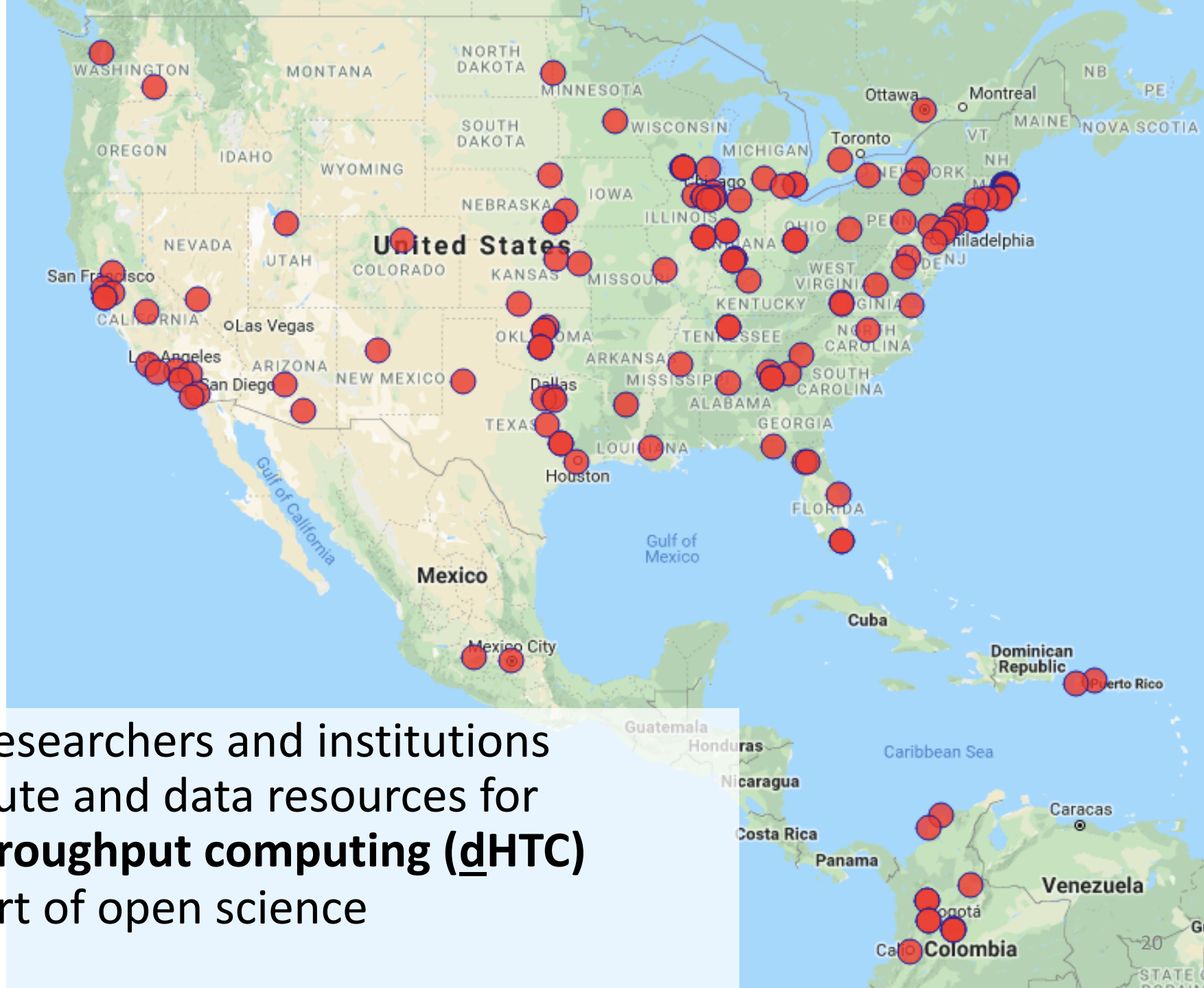
Then you …

> can reach your research goal after ***this much time and effort***, and

> could really achieve ***this much more research outcome.***

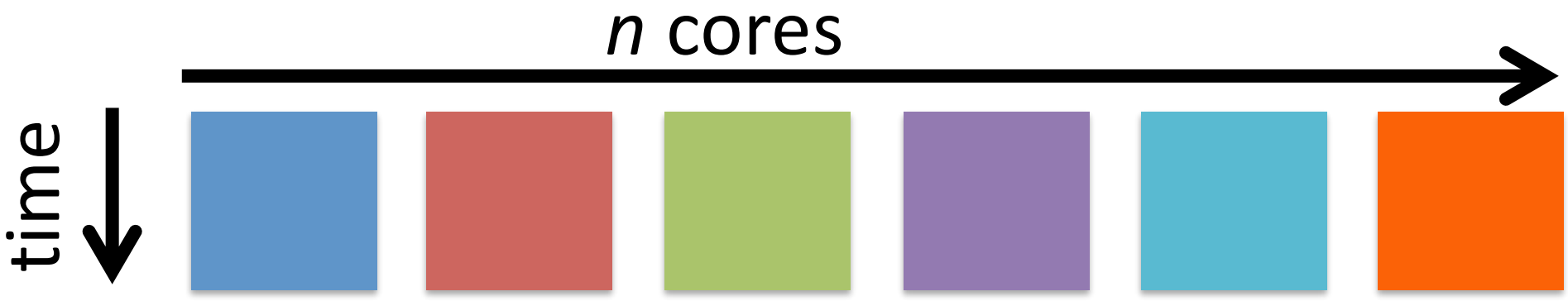
Open Science Grid

# What is the Open Science Grid?




a consortium of researchers and institutions who *share* compute and data resources for ***distributed* high-throughput computing (dHTC)** in support of open science

# distributed High Throughput Computing (dHTC)




*versus internal parallelism (multi-threading, MPI, etc.)*

- Scheduling: only need **1 CPU core, each** (shorter wait; faster to peak)
- No special programming required (laptop-style, no multi-threading, etc.)
- Easier recovery from failure (if one core fails, job requeued, other cores/jobs continue)
- Number of concurrently-running jobs is *more* important
- CPU speed and homogeneity are *less* important

# Is it OSG-able?

Is your work computable as lots of *laptop-sized* pieces?

Open Science Grid

| ***Per-Job*** **Resources** | **Ideal Jobs!** (up to 10,000 cores, per user!) | **Still Very Advantageous!** | **Probably not…** |
|---|---|---|---|
| **cores** (GPUs) | **1** (1; non-specific) | **<8** (1; specific GPU type) | **>8 (or MPI)** (multiple) |
| **Walltime** *(per job)* | **<10 hrs*** *or checkpointable | **<20 hrs*** *or checkpointable | **>20 hrs** |
| **RAM** *(per job)* | **<few GB** | **<10 GB** | **>10 GB** |
| **Input** *(per job)* | **<500 MB** | **<10 GB** | **>10 GB** |
| **Output** *(per job)* | **<1 GB** | **<10 GB** | **>10 GB** |
| *Software* | *'portable' (pre-compiled binaries, transferable, containerizable, etc.)* | *most other than →→→* | *licensed software; non-Linux* |

# HTC Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
3. **Set expectations.**

If you …

> execute your work ***this way*** (HTC)

> requiring ***these learning steps*** and ***this much work***.

Then you …

> can reach your research goal after ***this much time and effort***, and

> could really achieve ***this much more research outcome.***

# HTC Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
3. **Set expectations.**
4. **Follow up with the personalized plan for #3.**
5. **Make ongoing support <u>accessible</u>.**

***<u>Over</u>***-emphasize your willingness to help!

Open Science Grid

# HTC Facilitation Training

## *Shadow OSG's Research Computing Facilitators*

- **Join OSG RCFs in facilitating use of OSG Connect by researchers from *your* institution**
  - Get an account, yourself. Tell your researchers to include you in OSG support conversations.
- **on-site shadowing at UW-Madison**
  - We can include you in our researcher consultations for a week, with scheduled time for discussion/reflection.
- face-to-face community engagement at HTCondor and OSG events (PATh events)

# dHTC Training and Community Events for Campus Staff

*Learn dHTC skills and engage with others!*

- **Annual Community Meetings, including content specific to campus research:**
  - OSG All-Hands Meeting (~March)
  - HTCondor Week (~May)
- **Annual OSG User School**
  - materials online! (recently virtual)
- **Trainings, Workshops, and Presentations** at:
  - conferences, e.g. PEARC, CC*, RMACC, etc.
  - virtual communities, e.g. CaRCC, MAGIC, etc.

# OSG User School

- week-long summer education program on HTC and OSG compute execution
- targeted for researchers *and* research computing staff
- spring application/review
- all materials publicly available
  - 2019
  - virtual pilot in 2020

Open Science Grid

# Entry Points that YOU Control

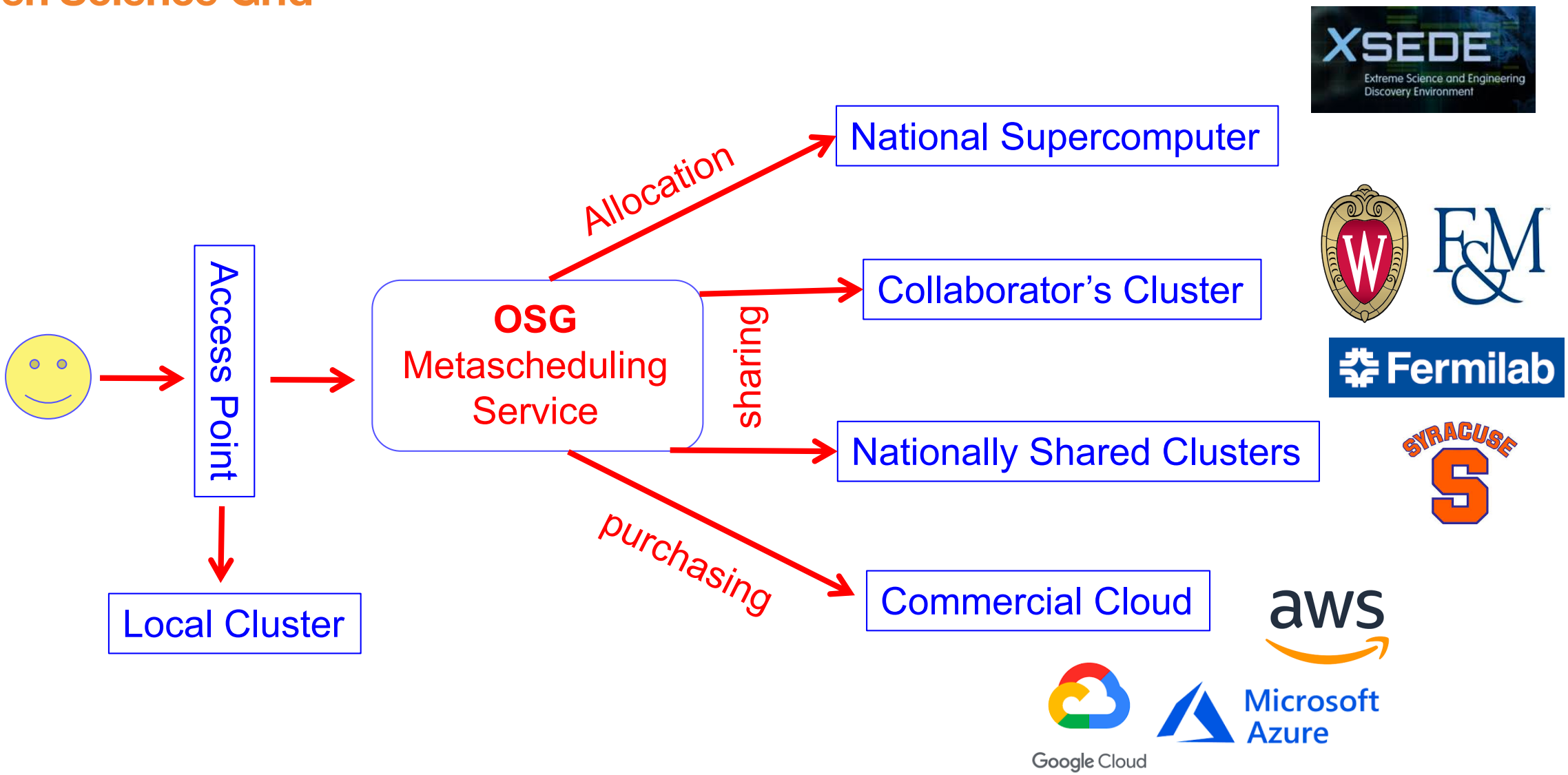
*… like having a cluster where you don't have to administer the 'worker' nodes, and can* ***provide all of the user support****.*

## The institution has primary responsibility for:

- user *facilitation*
  - incl. software portability (OSG modules available), troubleshooting
  - leverage support from OSG staff (co-facilitation, etc.)
- administering user authorization
- (some) HTCondor administration on the submit node
- administering/integrating any institutional data storage

***Resource sharing not required!***

# Services Overview



Open Science Grid

## dHTC and OSG Services for Campuses

- Build ‘Local’ dHTC Capacity with
- Share Local Resources via OSG
- Support Your Researchers to Use OSG
  - **TODAY! (via OSG Connect)**
  - dHTC and dHTC Facilitation training
  - Local submission points into OSG
- Support for CC* Proposals

***All “FREE” and OPEN, with Facilitation for Campuses***

Open Science Grid

Submit locally, run globally.

# Next Steps

- **You and Your Researchers Can Use OSG, NOW!**
  - **osgconnect.net > Sign Up**
- **Discuss dHTC Services for Your Campus**
  - **support@opensciencegrid.org**
- **Consulting and Letters of Support for CC* Proposals**
  - **cc-star-proposals@opensciencegrid.org** (to OSG leadership, only)