

Shared Responsibilities, Accounting, and Reporting

Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC





Shared Responsibilities



- **OSG provides:**
 - Scientific user communities to consume the 20% shared time.
 - We provide all the support to these scientists.
 - A mechanism for scientists to submit jobs to your cluster.
 - [Your cluster becomes part of the OSG Compute Federation Open Pool](#)
 - A uniform runtime environment, incl. data access, for these scientists.
 - Accounting data for your annual report to the NSF to show who benefited from your cluster, when, and how much.
- **You provide:**
 - A set of SSH-key accessible accounts from which we can submit jobs to your cluster.
 - Access to the uniform runtime environment of OSG.

([Technical details on both in Brian Bockelman's talk](#))
- **You decide: “when, who, and how much”**
 - Which of the communities that OSG supports gets how much access to your cluster and when.
 - You control this via your batch system configuration
 - You are entitled to kill any jobs any time you want
 - You are entitled to complain if we fail to use resources you make available.

Let me dive in a bit further ...

**We want to hear your input on anything that comes next.
If something is impractical to you, let us know.
We are always eager to improve our processes,
especially now as we start up PATH !!!**



Communications



- We need to be able to communicate with each other
 - We track contacts at your site, especially for security and system administration communications
 - We accept listserv if that's most useful to you.
 - When the names/emails of people change, we'd like to know within 2 business days.

- We'd like to know when things change on your side.
 - Announce planned cluster downtimes 2 business days in advance.
 - Announce unplanned downtimes, asap after they happen. Even if you recovered by the time you tell us.
 - Announce when things are back in order.
 - Announce any major changes:
 - OS, any changes in Queues, ...
 - Basically, anything you told us at the beginning, tell us when it changes.
 - We prefer you to “overshare” !
- We will let you know when things change on our side.
 - All of the above also applies in the inverse.

Does OSG have an MoU?

Wikipedia:

A **memorandum of understanding (MoU)** is a type of agreement between two ([bilateral](#)) or more ([multilateral](#)) parties. It expresses a convergence of will between the parties, indicating an intended common line of action.^[1] It is often used either in cases where parties do not imply a legal commitment or in situations where the parties cannot create a legally enforceable agreement. It is a more formal alternative to a [gentlemen's agreement](#).^{[2][3]}

- No ... at least not yet.
- We are drafting something we'd like to call "Policy on Shared Responsibilities"
- In addition to the previous slides, it will contain:
 - Pointers to other policies, like:
 - Acceptable Use Policy
 - Security policies and expectations
 - Policy on accounts, access, privacy
 - OSG Principles of Sharing

While we have these today, we want to spend some time to review them in light of starting up PATH.

Accounting and Reporting

We are here to help you provide reports to not just the NSF, but also any other entities you need to report to.

It is not our role to tell you what to report to whom.

We simply provide information.

You choose how to use it !!!

- GRACC accounting
 - Brian B. talked about this already
- Project registration in github
 - Each account that has access to the Open Pool must register with a project.
 - Each project has a minimal description registered with us.
 - You are welcome to use those descriptions for your reports.



Example projects descriptions



Department: Particle Physics

Description: High intensity frontier experiment searching for physics beyond the Standard Model

Model

FieldOfScience: High Energy Physics

ID: '346'

Name: REDTOP

Organization: Fermilab

PIName: Corrado Gatto

REDTOP

CatalystHTVS

Name: CatalystHTVS

Description: Using high throughput computing to screen molecular catalysts for energy fuel conversion based on experimental database or in-silico generated structures. In the next stage, the output from HTC calculations will be used to train machine learning models to allow faster and higher throughput molecular catalyst design.

Department: Chemical Engineering

FieldOfScience: Physical Chemistry

Organization: Massachusetts Institute of Technology

PIName: Heather J. Kulik

Department: Neurological Surgery

Description: 'The hope for magnetoencephalographic (MEG) measurements has been to produce functional brain mapping with high spatial (mm) and temporal (msec) resolution. Realizing this hope requires answers to these questions: (1) How many sources are active within the brain? (2) Where are they located. (3) What is their time course? MEG Virtual Recording (MVR) provides these while producing noninvasive measures of intracranial neuroelectric currents as if from 2,000,000+ directly implanted electrodes. It does so from single trial (unaveraged) data, has no free parameters, and provides very strong probabilistic measures to validate the existence of each identified source. We have demonstrated efficient implementation of MVR on the Open Sciences Grid. The measured computational load of 400 SU per second of MEG data makes supercomputing essential to practical implementation of MVR. We anticipate that MVR will enable identification of specific neurophysiological biomarkers of a variety of non-structural brain pathologies which have been refractory to date, e.g. concussion, post-traumatic stress disorder.'

FieldOfScience: Biological Sciences

ID: '33'

TG-IBN-13001

Name: TG-IBN130001

Organization: University of Pittsburgh

PIName: Donald Krieger

**You have seen the accounting we have
in Brian B.'s talk earlier today.**

Next, I'll show you what I'd like us to have for you.
I'll use the example of our covid19 accounting page.



Open Science Grid

Covid19 Accounting in GRACC



COVID-19 Research



Last 1 year UTC



COVID-19 Research on the Open Science Grid

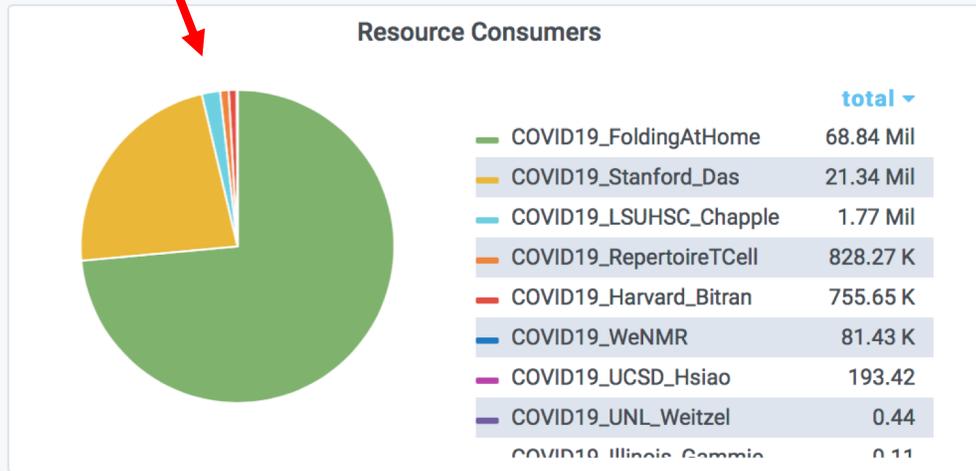
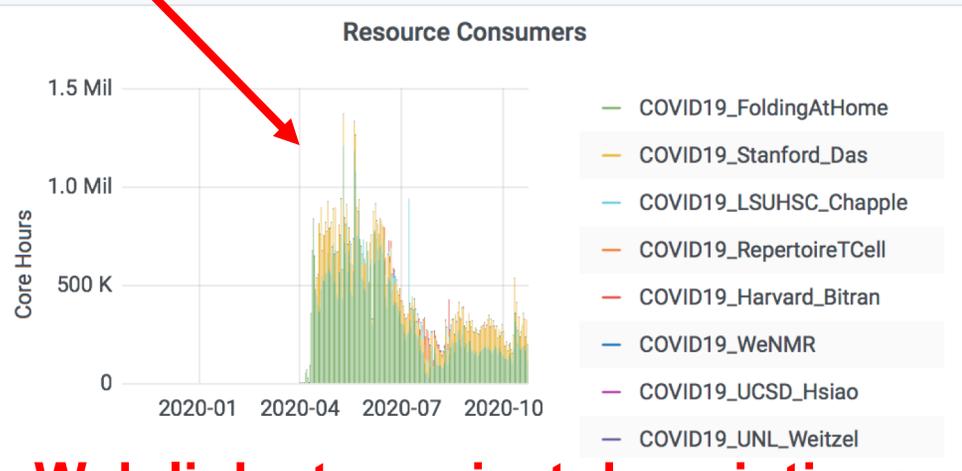
Researchers and resource providers can utilize the Open Science Grid.

- [Resource Provider Documentation](#)

“Who” integrated over time window

“Who” versus time

Who is performing COVID-19 Research?



Web links to project descriptions

List of Projects

Project Name	PI Name	Institution	Core Hours Consumed
COVID19_Stanford_Das	vinja Das	Stanford University	21.34 Mil
COVID19_LSUHSC_Chapple	Andrew Chapple	LSU School of Public Health	1.77 Mil
COVID19_RepertoireTCell	Michael Noble	Repertoire Immune Medicines	828.27 K

Following the web link to project description

28 lines (26 sloc) | 1.55 KB

```
1 Name: COVID19_LSUHSC_Chapple
2 Description: "Most Bayesian methods require Markov Chain Monte Carlo sampling (MCMC)
3 to obtain posterior distributions, which can be used for statistical inference
4 - and decision making during adaptive clinical trial designs. To justify any novel
5 statistical method or adaptive design, extensive simulation studies must be
6 conducted to demonstrate their effectiveness. Dr. Chapple recently used OSG to
7 successfully revise a novel statistical method for survival analysis relevant to COVID-19.
8 Such simulations, particularly for Bayesian adaptive clinical trials,
9 can take a tremendous amount of time to run 1,000 or more simulations for a given
10 scenario, and usually hundreds of scenarios are warranted to convince others of the
11 trial's benefit. Dr. Chapple has used 435 thousand core hours to develop clinical
12 trial designs for testing safety of new agents in pediatric brain tumors, testing
13 multiple COVID-19 therapies simultaneously, and determining optimal treatments
14 based on patient subgroups. Without OSG, it would not have been possible to start
15 enrolling patients in a 3-treatment armed COVID-19 trial at University Medical
16 Center in New Orleans, LA. Based upon that success, Chapple will also demonstrate
17 the same approach for 5 treatment arms, and also for subgroups (based on
18 comorbidities, age, etc), and publish the trial design in a statistical journal."
19 Department: Biostatistics
20 FieldOfScience: Health
21 Organization: LSU School of Public Health
22 PIName: Andrew Chapple
23
24 ID: '671'
25
26 Sponsor:
27 CampusGrid:
28 Name: OSG Connect
```



In addition ...

- Site will be selectable just like Brian showed earlier today.
- Consumption vs time
- Consumption in aggregate for chosen time window
- By project and by field of science and by institution of the PI of the project
- **And anything else you'd like to see.**

**Let us know how our accounting
can be more useful to you.**



Summary & Conclusion

- OSG's objective is to “Advance Open Science through distributed High Throughput Computing”
- We think of ourselves as a part social and part technical project.
- To achieve our objective, we want to engage with you, and build long term relationships.

Thanks !

Contact us at: help@opensciencegrid.org