



Building communities via the Open Science Grid

**Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC**

A bit of History

- The OSG was created in 2004 as a collaboration between:
 - Science Collaborations
 - Technology Providers
 - Resource Providers
- with the goal to “*design, deploy, and operate a shared distributed infrastructure and services in support of US science*”.

(excerpt from OSG Management Plan)

Charter of the Open Science Grid

Draft V0,1, October 2004

Open Science Grid Governance Technical Group

Increasing scale and complexity of 21st century science has led to larger and more global collaborations involving massive data sets. This requires efficient utilization of widely distributed computational resources and effective global communication. Among the sciences that can immediately benefit from global computing grids are high energy physics, nuclear physics, astrophysics, computational biology, and more.

**The full document is just about 1 page long,
and after 12+ years still
surprisingly relevant to what we do in practice.**

Increasing scale and complexity of 21st century science has led to larger and more global collaborations involving massive data sets. This requires efficient utilization of widely distributed computational resources and effective global communication. Among the sciences that can immediately benefit from global computing grids are high energy physics, nuclear physics, astrophysics, computational biology, and more.

The vision of the Open Science Grid Consortium (OSG) is one of a persistent production national grid infrastructure for large scale US science: the Open Science Grid.

The Open Science Grid Consortium will provide a set of goals and an overall infrastructure within which the Grid resources of the different members can be operated coherently and compatibly.

The Open Science Grid Consortium is a consortium of scientific collaborations, scientific computing centers and existing and new grid research and deployment projects, involving both computational and application scientists, working together to provide and support the set of facilities, services and infrastructure needed.

The Open Science Grid Consortium will require a dedicated staff and other resources, which could be paid for by member contributions or eventually by direct funding.

A structure of management and coordination bodies will oversee and coordinate the work of the Open Science Grid Consortium.

The Open Science Grid will be open to all sciences that have a need for distributed large scale computing and data management, and can bring resources to be federated.

An engineered production quality grid infrastructure will be built and operated in the US, and extended internationally to participate in the global grid infrastructure for science. To meet the data analysis needs of the scientific communities the grid will support managed access to 10s of thousands of computers and 100s of petabytes of storage.

The Open Science Grid will create opportunities for educators and students to participate in building and exploiting this grid infrastructure and opportunities for developing and training a scientific and technical workforce. It has potential to transform the integration of education and research at all levels.

The Open Science Grid Consortium will ensure that the U.S. plays a leading role in defining and operating the global grid infrastructure needed for large-scale collaborative and international scientific research. The Open Science Grid will provide a set of services that can be enriched as new science areas choose to join and federate their resources. For the first time combined computing resources at several national labs and at dozens of universities will effectively become a single national computing infrastructure for science, the Open Science Grid.

It's been revised in 2006 and 2011.
All versions are available online as
Document 25 in OSG Docdb.

<http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=25>

- OSG Council
 - unfunded governance body to chart the direction of the consortium, and elect the Council Chair and Executive Director (ED). For more details google “OSG Council”.
- OSG Project led by the Executive Team (ET)
 - funded project that runs the day-to-day operations of the infrastructure and services. ET is lead by ED.
- OSG Consortium
 - anybody that uses or contributes to the software, infrastructure, and/or services.

Relationship between these three are determined by governance documents and 10+ years of practice

Software & Technology

incl. compute, storage, networking

Operational Services

Community Engagement

incl. user & admin support, training

Goals & plan of work is set annually based on stakeholder input.

Area coordinators oversee execution of plan.

Course corrections can occur any time in response to tickets & requests.

Anybody in the consortium can influence the plan of work.

The plan is managed by ED supervised by council.



Vision of OSG

Research Computing is the new Library

- Over hundreds of years, the defining common research service at Universities was the Library.
 - defining service was the curation of information to support the creation of knowledge
- Modern Science needs so much more ...
 - compute, storage, networking, ...

=> Cyberinfrastructure

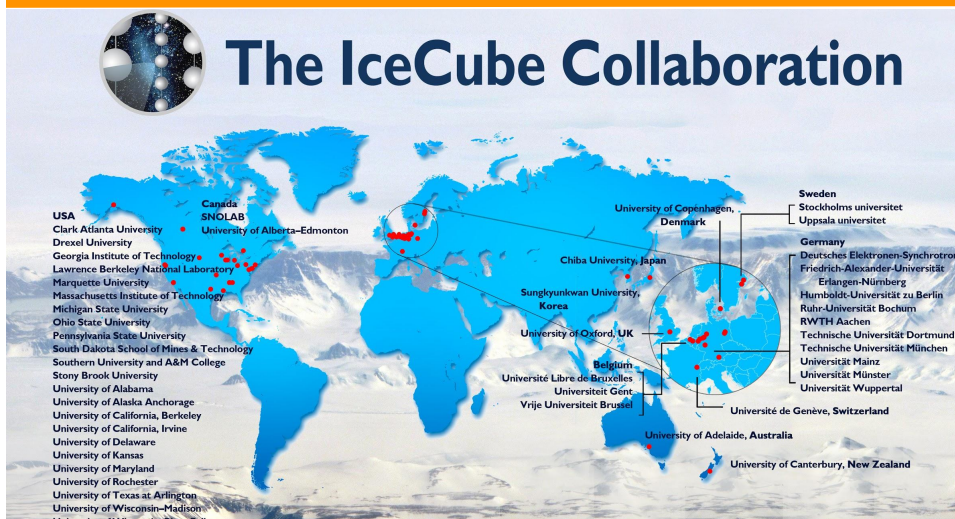


Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked by high speed networks to make possible scholarly innovation and discoveries not otherwise possible.

Indiana University Knowledge Base
... found by fkw via google.

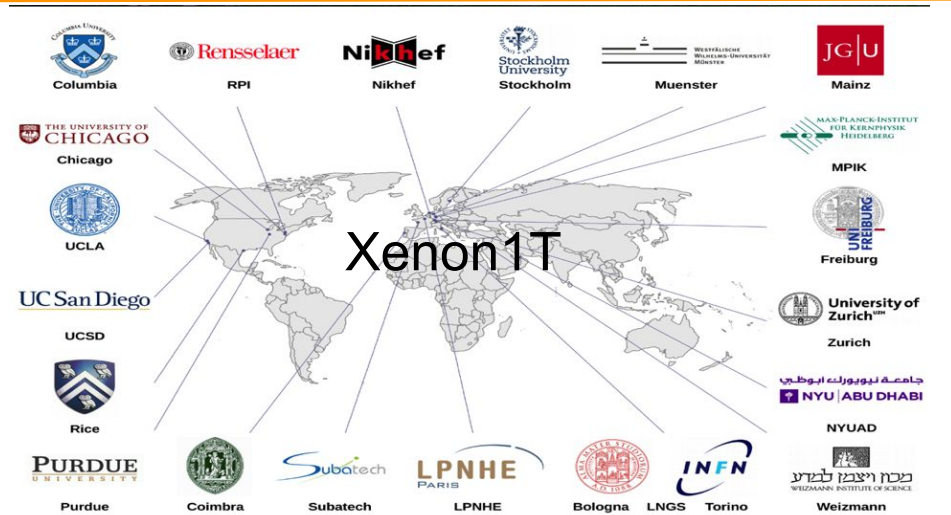
To advance Open Science, Universities will increasingly need to provide Cyberinfrastructure as a common good for their scientific communities.

Science is a Team Sport



- ~100 members, 20 institutions
- 24 non-affiliated members
- +35 associate members
- Smithsonian Astrophysical Observatory
- Adler Planetarium
- Argonne National Lab
- Barnard College / Columbia University
- Bartol Research Institute / University of Delaware

- Georgia Institute of Technology
- Iowa State University
- Purdue University
- University of California, Los Angeles
- University of California, Santa Cruz
- University of Chicago
- University of Iowa
- University of Minnesota



- University of Utah
- Washington University in St. Louis
- McGill University, Montreal
- University College Dublin
- Cork Institute of Technology
- Galway-Mayo Institute of Technology
- National University of Ireland, Galway

VERITAS



SPT3G



Open Science Grid

An international team sport



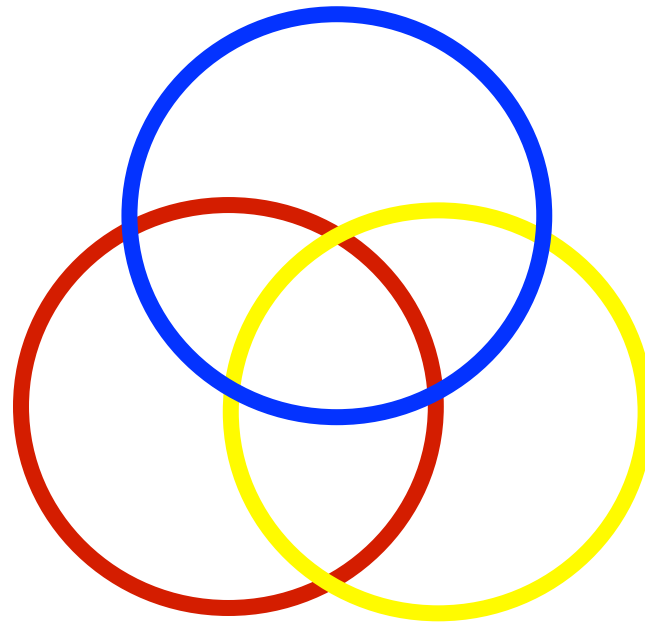
**ATLAS or CMS include
~ 200 institutions across ~ 40 countries**



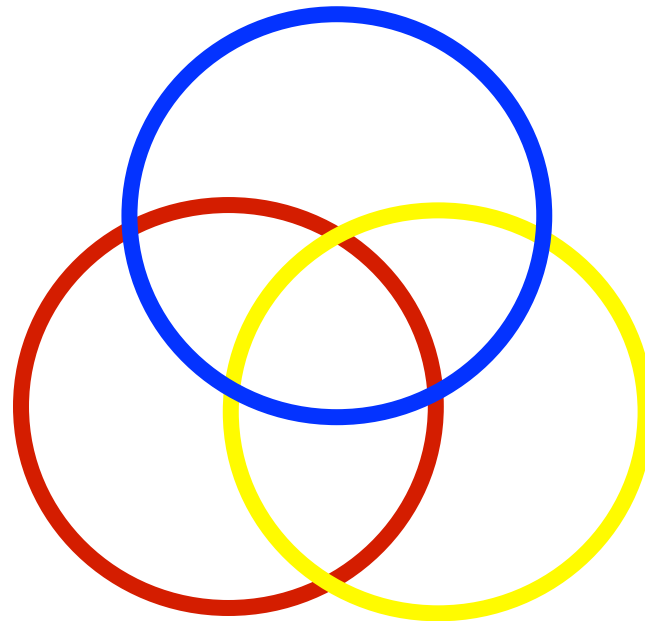
OSG advances the local, national and international **integration of Cyberinfrastructure **in support of Open Science.****

Even “more moderate size” physics experiments than LIGO, ATLAS, or CMS involve **dozens of institutions across multiple countries** that need to be able to **share their resources** to **maximize their scientific throughput !!!**

We see this phenomenon of multi-institutional teams as a striking commonality across research disciplines.



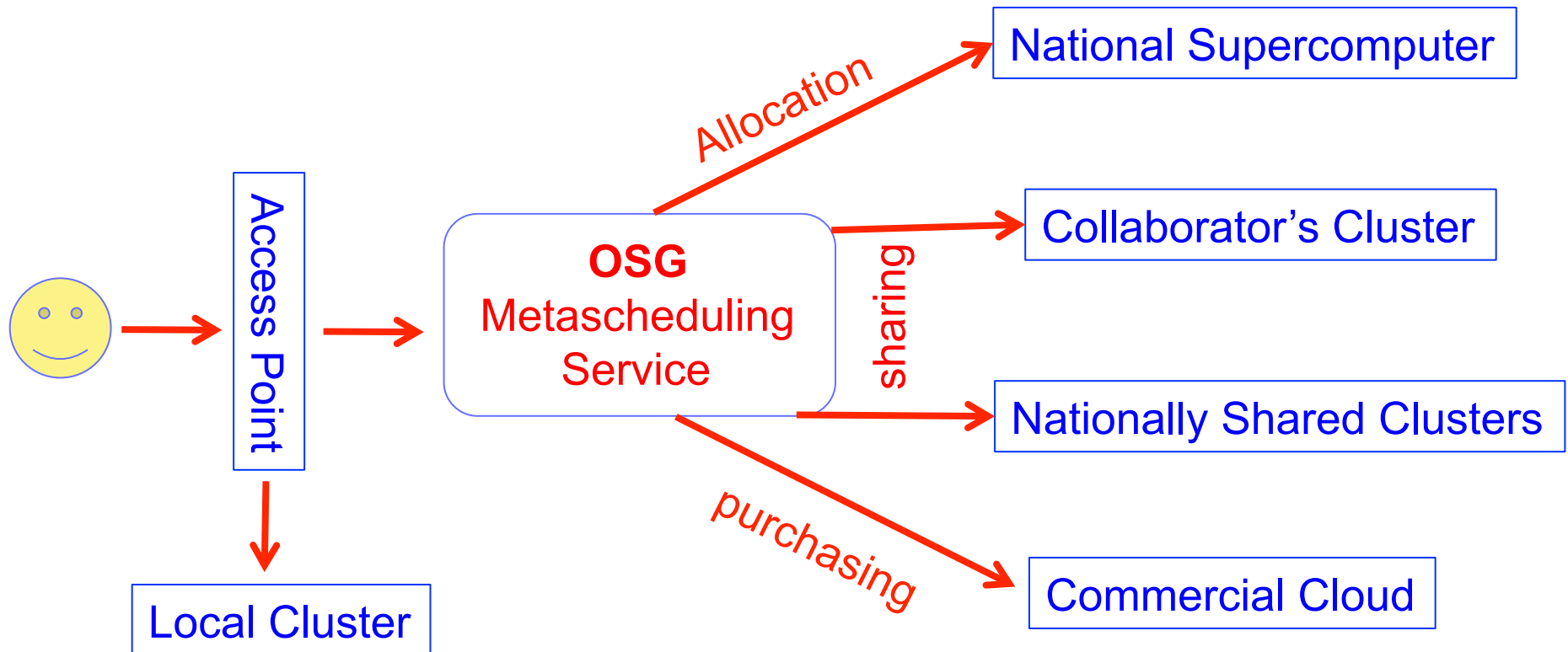
because they occupy the overlap
in the venn diagram of the science teams



It is not uncommon for OSG to introduce scientists to each other, or for scientists to refer colleagues to OSG.

Integrating CI across campuses means integrating CI for many science teams at once !!!

Transparent Computing across different resource types



OSG integrates computing across different resource types and business models.

Everybody gains from the multi-disciplinary focus of OSG

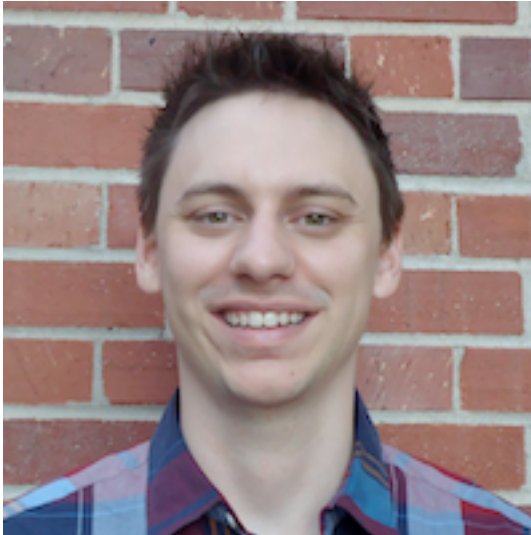
Given the attention to Machine Learning at this workshop, we'll use TensorFlow as example for how the introduction of a new capability to OSG benefitted from the multi-disciplinary focus of OSG.

The Science Use Case

- A post-doc in biomedical informatics at Columbia Medical School wants to use TensorFlow to mine data in FDA reports for single drug effects as well as drug interactions.
- This requires many thousands of GPU hours.
- He's familiar with HTCondor and OSG from a previous published science project, and has a GPU development platform, and application ready to go.
- He's willing to bleed a little at the bleeding edge in order to help OSG develop a new capability.

- IceCube and SBGrid (Structural Biology) already pioneered the use of GPUs on OSG last year.
- UNL allows opportunistic access to their GPUs.
- OSG technology group introduced singularity containers within the last 6 months
 - today 25-75% of opportunistic use of OSG is using singularity under the hood without the users being aware of that.

OSG Team to make it work



Derek, Computer Science



Bala, Condensed Matter Theory



Mats, Computer Science



Brian, Math & CS

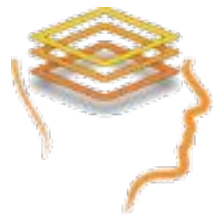
While both the resources on and the use of OSG is obviously dominated by the LHC community, the people on the OSG project have a much broader background.

With the advent of ML@OSG we expect increasingly small clusters of cheap GPUs to be added to support it.

This is driven by science other than the LHC.

Summary & Conclusion

- OSG has fostered multi-institutional and multi-disciplinary collaboration since 2004.
- We support science at all scales, from single PI to large international collaborations.
- We support infrastructure at all scales from small lab clusters to supercomputer centers to commercial clouds.
- Anybody working with OSG decides for themselves who may be part of their community.
 - there are examples of institutions that only support a specific subset of science communities on their resources.



opensciencegrid



[osgusers](#)



user-support@opensciencegrid.org



support.opensciencegrid.org



www.opensciencegrid.org/links