Open Science Grid

Fermilab

The OSG Open Facility: A sharing ecosystem

Bodhitha Jayatilaka, Tanya Levshina, Chander Sehgal, Marko Slyz

Fermilab

Mats Rynge

ISI/USC

CHEP 2015

Okinawa, Japan

April 13, 2015

The Open Science Grid

- Goal: **Advance science through open distributed high throughput computing** (DHTC) in the US
  - Operates as a partnership between **resource providers** (sites) and **stakeholders** (scientific communities of users)
  - OSG **project** provides middleware and R&D effort
  - Over 120 sites (computing and storage elements)
    - Mix of university and national labs; LHC experiment computing sites and campus clusters
  - Uses the **Virtual Organization** (VO) model
    - Most VOs correspond to large HEP experiments and university communities
- Funded by the **US Dept. of Energy** and **National Science Foundation**
  - Second 5 year period of support

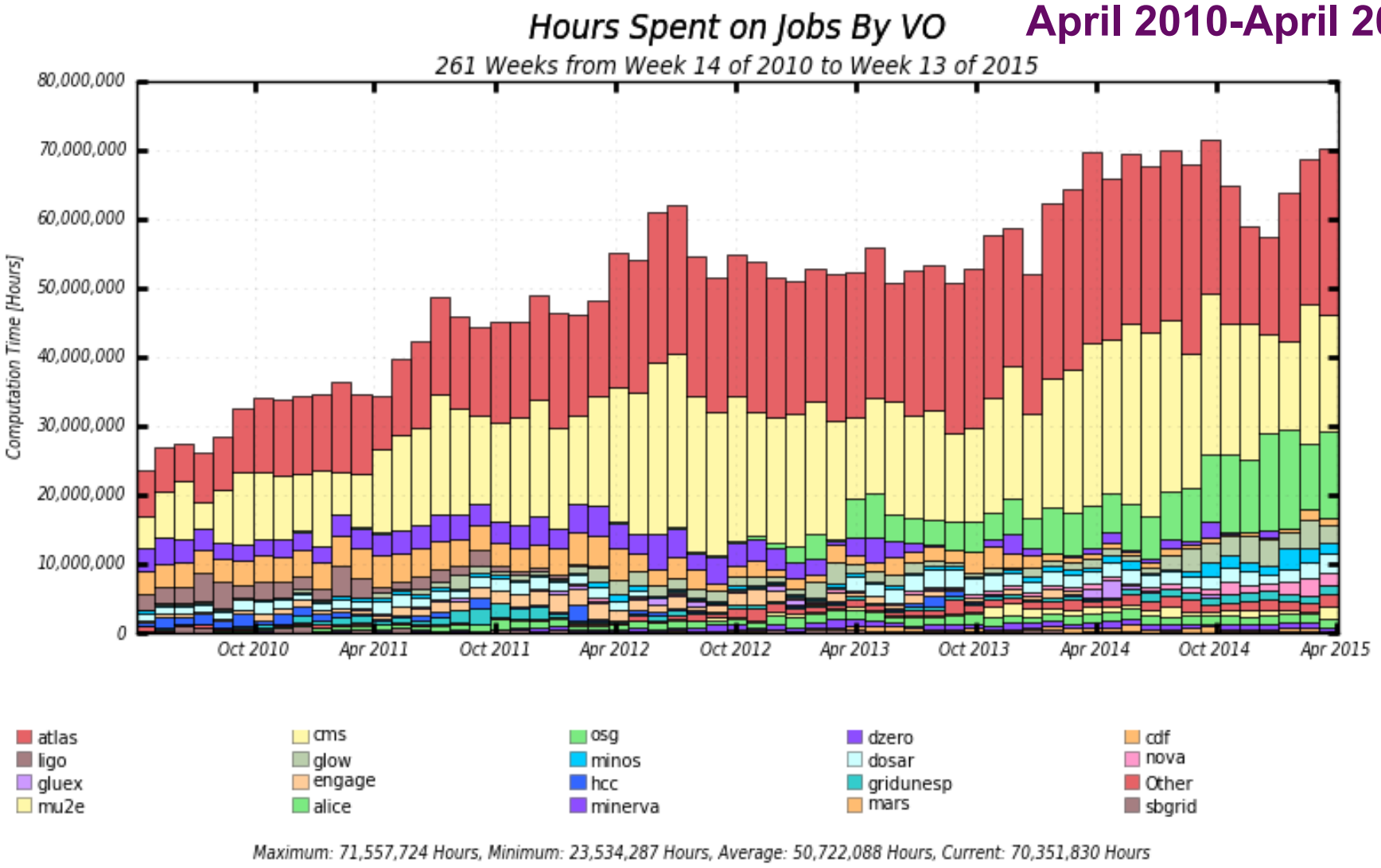
State of the OSG

In the last 24 Hours	
664,000	Jobs
1,633,000	CPU Hours
2,423,000	Transfers
330	TB Transfers
In the last 30 Days	
15,248,000	Jobs
69,955,000	CPU Hours
74,181,000	Transfers
13,151	TB Transfers
In the last 12 Months	
196,761,000	Jobs
808,076,000	CPU Hours
587,639,000	Transfers
157,000	TB Transfers

http://display.grid.iu.edu

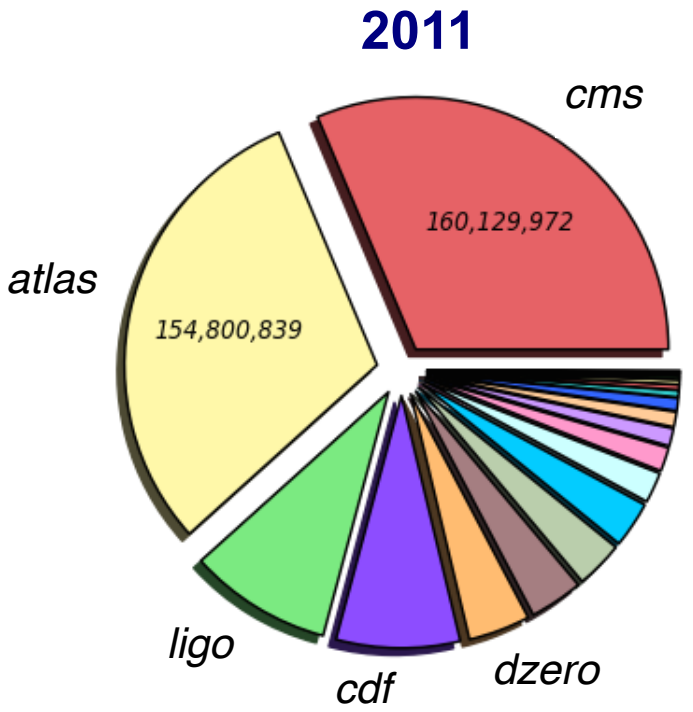
Growth of the OSG

April 2010-April 2015



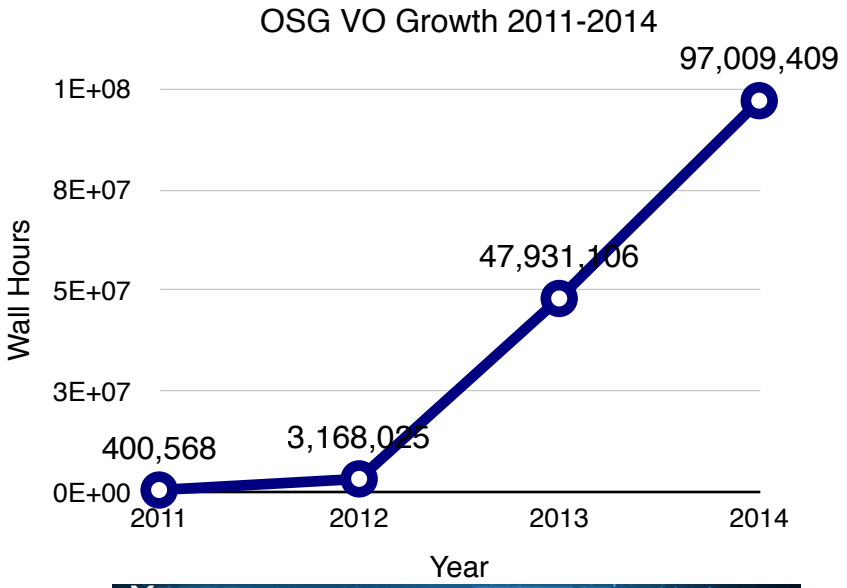
Expanding usage of the OSG

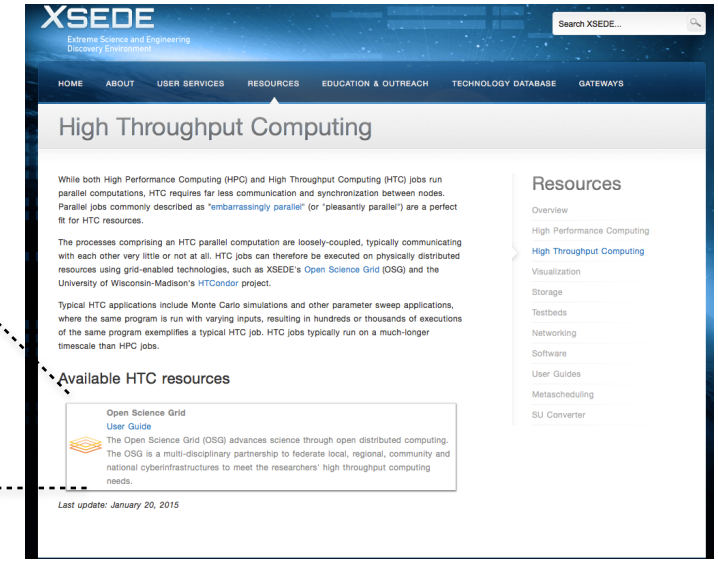
- Most use of the OSG is by large physics experiments (*e.g.* ATLAS and CMS) and individual university communities
  - Primarily operate on their own sites
    - Utilization is high but not 100% 24/7
  - 85% of OSG hours in 2011 were from large HEP experiments
- Want to connect unused cycles with researchers who are not part of large VOs
- Reaching other sciences takes more work
  - Few have the organized computing efforts HEP experiments have
  - Not every research community can have their own VO (although many do)



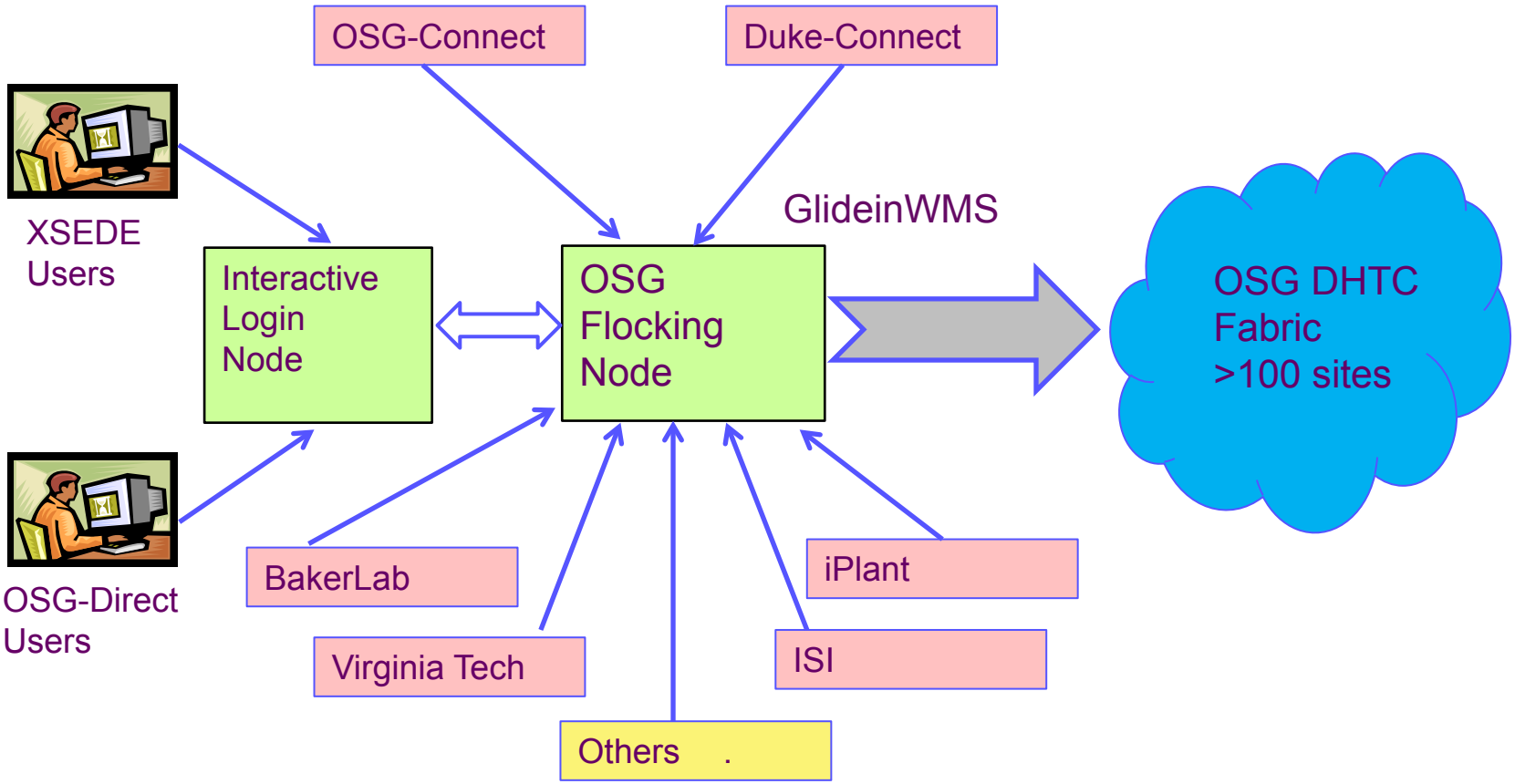
The OSG open facility

- Implement **OSG VO** as a VO for individual researchers
  - Allow **any researcher** from US institutions
  - Resources accessed by the OSG VO are **entirely opportunistic**
  - All OSG sites encouraged to support OSG VO running
- An **XSEDE** service provider
  - Currently the only DHTC provider

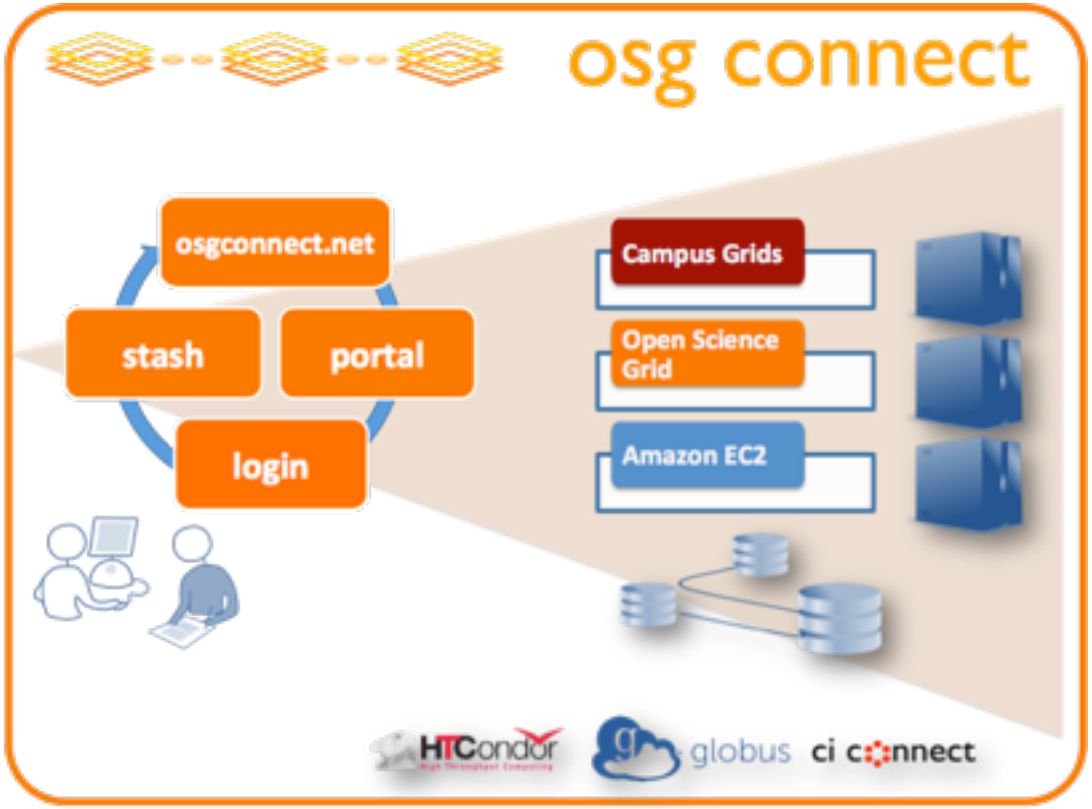




On-ramp to the open facility

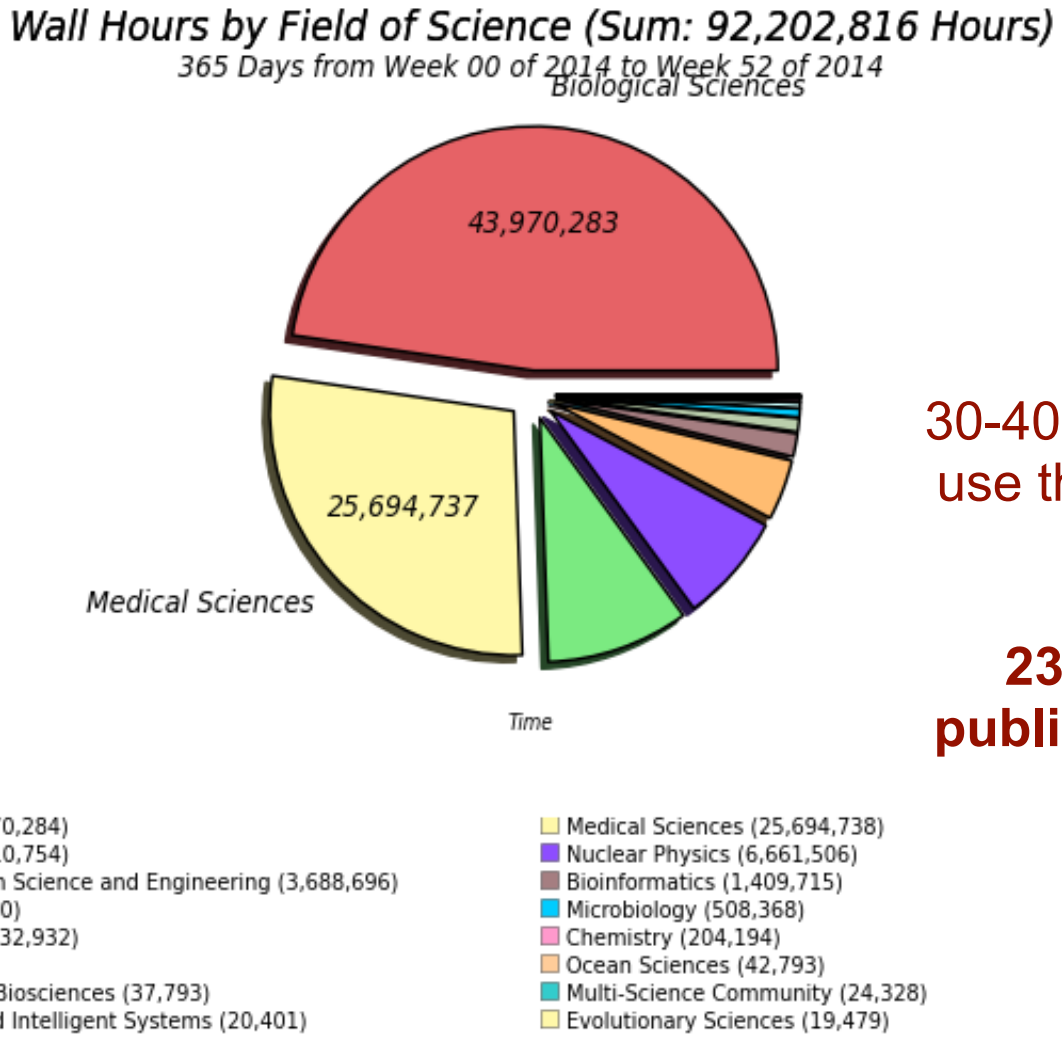


OSG Connect



- Make OSG seem like a **virtual campus cluster**
  - Login host
  - Environmental modules
  - Job scheduling
  - Job data (including monitoring)
- Jobs run on the OSG as part of the OSG VO

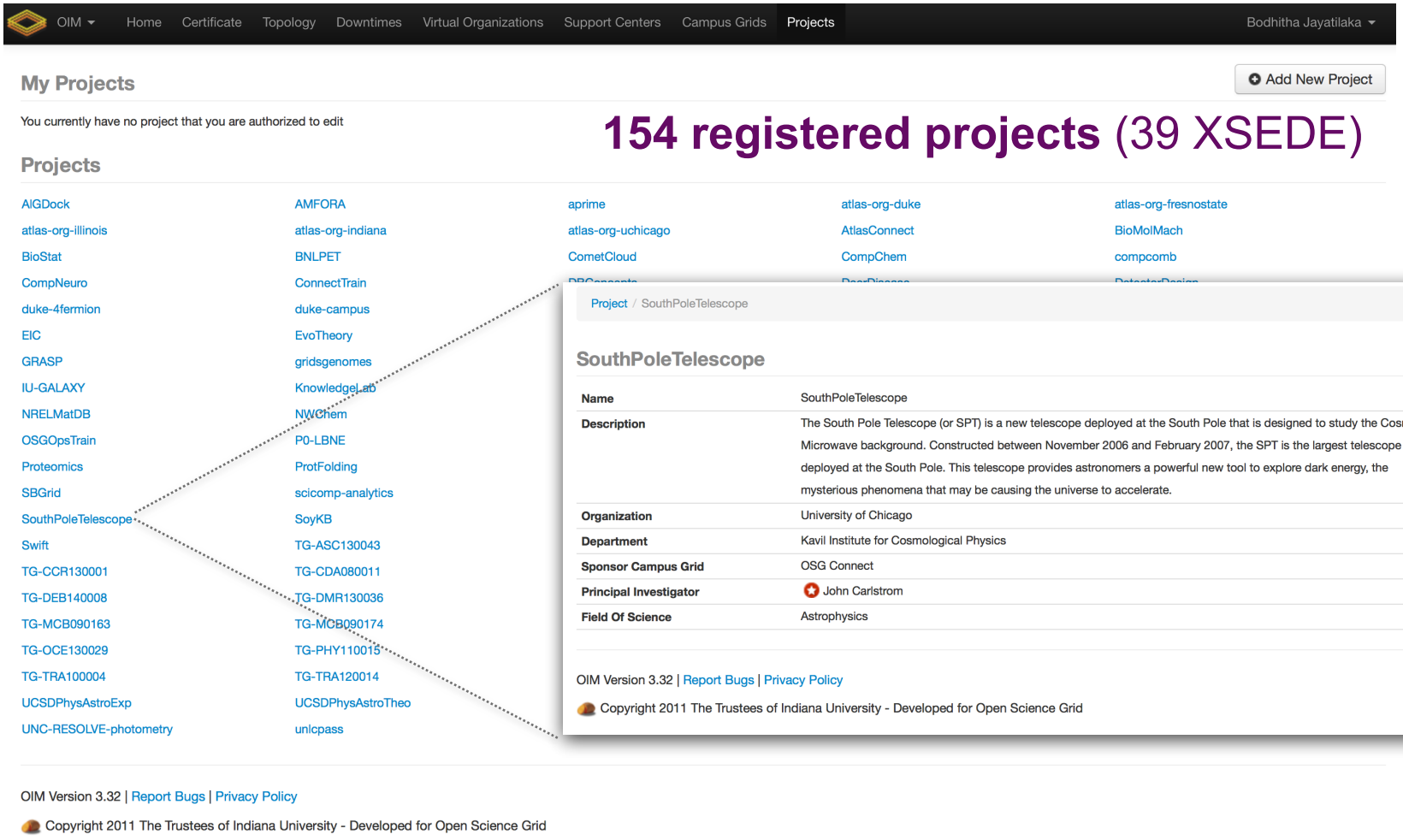
Science on the open facility



30-40 unique projects use the Open Facility every month

23 peer-reviewed publications in 2014

Projects using the open facility



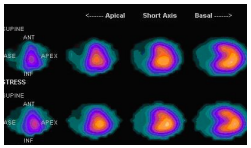
https://oim.grid.iu.edu/oim/project

Science on the open facility

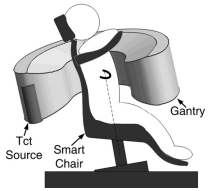
Detector Design

John Strologas, Wei Chang
Rush University

- Non-invasive diagnosis of coronary artery disease requires imaging of the myocardium
  - Predominantly with SPECT (single photon emission computed tomography)
  - Radio-pharmaceuticals+collimators and radiation detectors allow 3D imaging of the myocardium
  - Literally see areas not getting enough blood



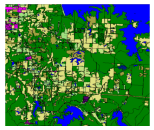
- Design of new C-SPECT optimized for cardiac imaging (cheaper x2 and more sensitive x2-3)
  - Extensive Monte Carlo simulations utilized in design studies as well as in calibrating prototypes

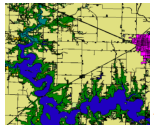


Agent-based modeling of disease transmission in white-tailed deer

Lene Jung Kjaer
Southern Illinois University

- Chronic wasting disease (CWD) in white-tailed deer
  - Prion disease (like mad cow) causing fatal degeneration of the brain
  - Not much known about the disease and its long-term effects on deer population
- Simulate spread of CWD with an agent-based model (DeerLandscapeDisease)
  - Each deer modeled separately with its own behavioral rules (landscapes, empirically-based movement, deer behavior, disease)
  - OSG used for CPU
- DLD shows transmission may be a mix of direct (contact) and indirect (environmental) transmission
  - Potential for coexistence
  - Landscape and social structure play a major part in CWD transmission

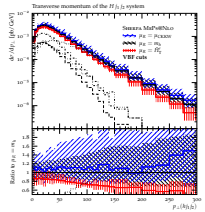


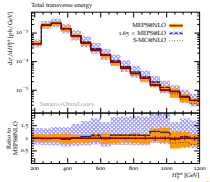


Collider phenomenology on the grid

Stefan Höche
SLAC

- LHC final states are very complex and interesting signals are often in corners of phase space
- Precise theoretical prediction of known processes crucial in finding new physics
- OSG resources used for refining modeling of SM processes and reducing theory uncertainties
  - gluon fusion Higgs+jets production
  - top quark pairs+jets
  - Both cases ≤2 jets@NLO and 3 jets @LO
- Uses Sherpa+Rivet+MCFM+OpenLoops
- Each process used O(500k) hours over 1-2 weeks

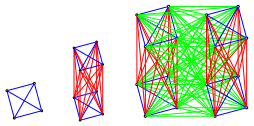


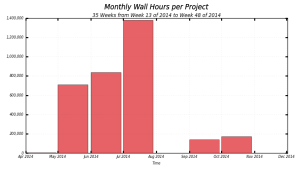


Enumeration of all ($C_5,C_5,C_5;n$) Colorings

David Narváez,
Stanislaw Radziszowski
Rochester Institute of Technology

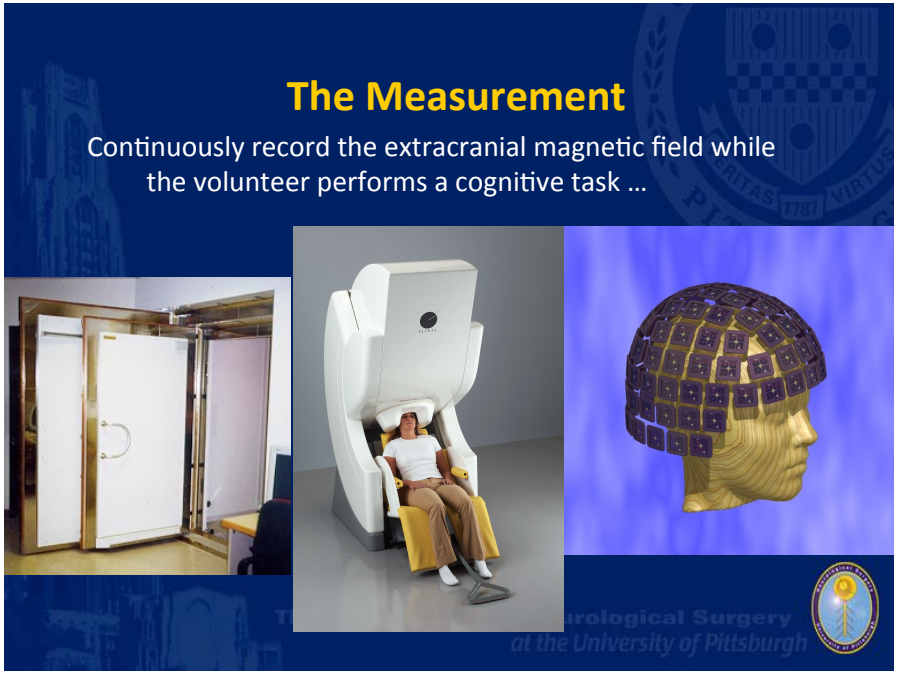
- A ($G_1,G_2,...,G_k;n$) coloring is a coloring of the edges of the complete graph on n vertices with k colors such that the i^{th} color does not contain the graph G_i
- $R_3(C_5)$=17. How many (C_5,C_5,C_5;16) colorings are there?
- Start with smaller C_5 colorings
  - Initial input of 140M n=12 colorings
  - Iterated by adding one vertex at a time
- Result: 1,701,746,176 Ramsey 3-colorings avoiding C5, the cycle of length 5, in all 3 colors





"Science on the OSG in 2014"
OSG AHM 3/25/15

Science on the open facility

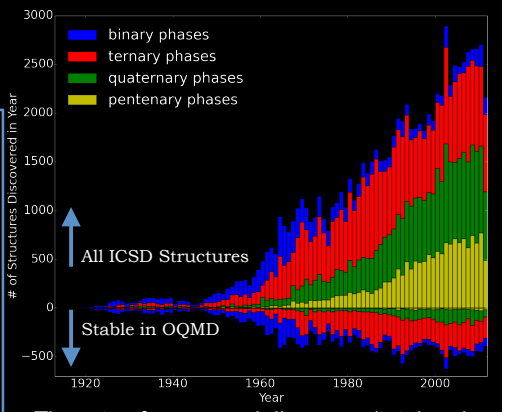


Combatting Concussions using "Virtual Recordings" (MEG)
D. Krieger (Pittsburgh)

Building the Open Quantum Materials Database
C. Wolverton (Northwestern)

How many compounds in our database are stable?

OQMD database:
- Total 297,099 compounds
- 19,757 T=0K stable
- 16,118 from ICSD
- 3487 "prototype" structures



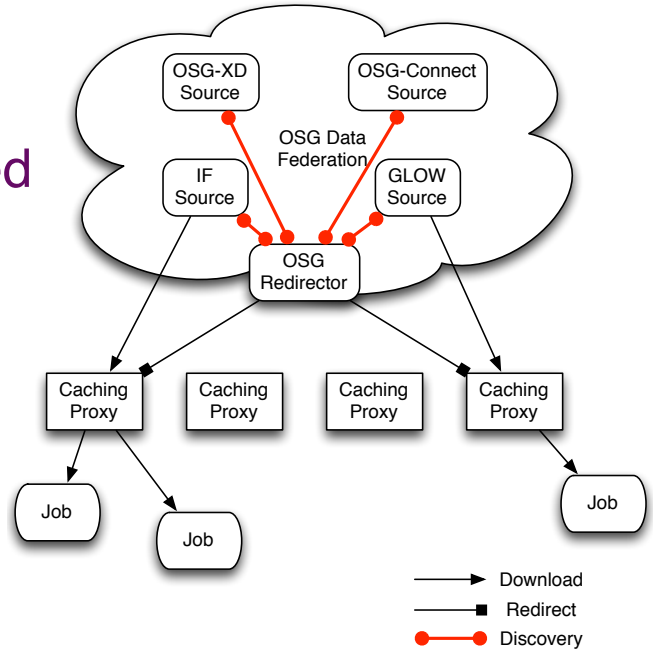
The rate of compound discovery (total and stable) within the ICSD by year.

Each of these cases represents a prediction of a system where new compounds should exist! Many gaps in our current knowledge of phase stability…

Future growth and technologies

- Open Facility now provides more than 100M CPU hours per year
- Ongoing challenges
  - Understanding available resources after LHC Run 2 starts
  - Finding appropriate users to provide continued demand
- Developing auxiliary technologies useful for opportunistic users
  - **Data movement** is a key limitation
  - Currently inefficient beyond ~10GB/job
- Developing **XRootD**-based caching system "StashCache"
  - Increases effective data movement to ~1TB/job

StashCache



Open Science Grid

Conclusions

- Opportunistic resources and their consumption continue to grow on the OSG
  - OSG Open Facility provides more than **100M CPU hours/year**
- A vibrant opportunistic user community co-exists with large (site-owning) VOs
  - A wide range of research is conducted (and published) using results obtained using OSG resources
- OSG Consortium committed to supporting these users in addition to large VOs
  - Principle of sharing is key to OSG vision and goals
  - Enshrined in by-laws: *Consortium members recognize that the OSG is a sharing eco-system and **strive to maximize** the sharing of computing resources, software, and other assets to enable science.*

Backup