



# Open Science Grid



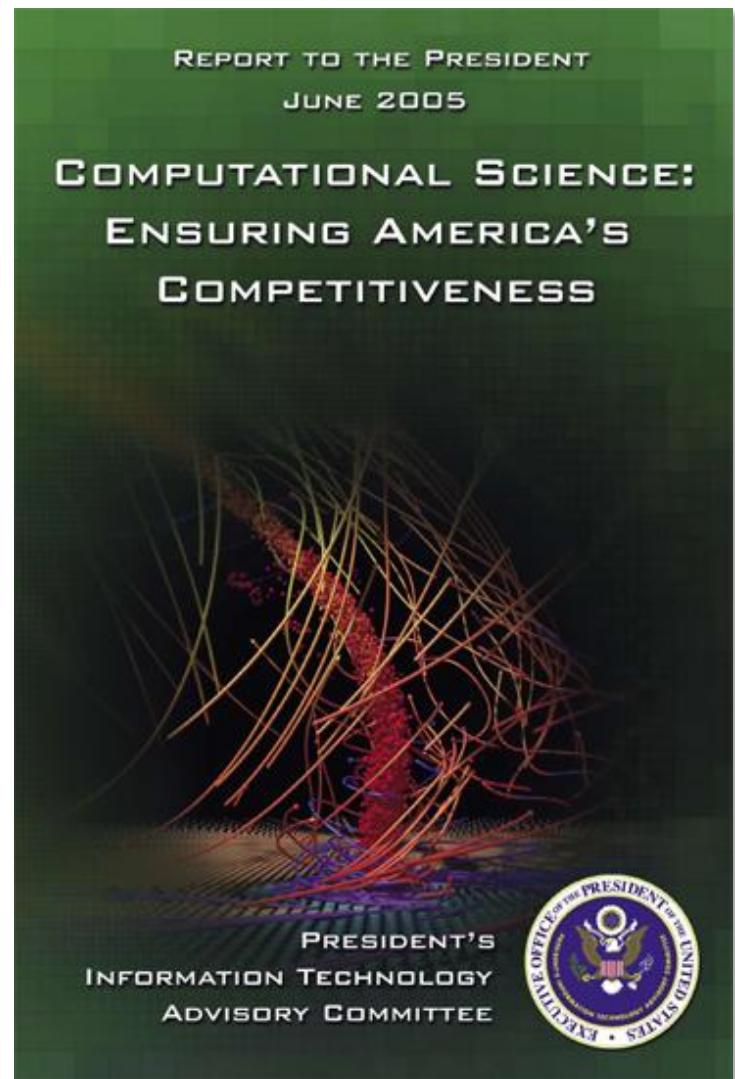
**John McGee – [mcgee@renci.org](mailto:mcgee@renci.org)**

Renaissance Computing Institute  
University of North Carolina, Chapel Hill

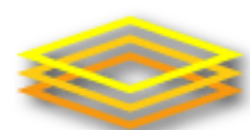




**Computational Science is a  
critical component of  
discovery, and its  
importance is growing**



[http://www.nitrd.gov/pitac/reports/20050609\\_computational/computational.pdf](http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf)



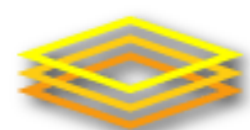
## HTC vs. HPC

- *High Throughput Computing (HTC)*
  - simple serial jobs, monte carlo, parameter sweep
  - complex workflows with data and/or job dependencies
  - ensembles of small node count MPI jobs
  - must be able to express the work in a highly portable way
  - cannot care about exactly where pieces are executed
  - do not have SSH access to resources; many implications
  - must be able to respect the community rules of a shared infrastructure; “plays well with others”
  - remarkable scaling opportunities



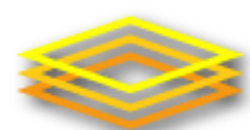
## HTC vs. HPC

- *High Performance Computing (HPC)*
  - top N computational scientists in the world
  - many important problems simply cannot be solved any other way
  - what does it take to successfully develop or run models at 256, or 512+ node parallelism?
  - what percentage of researchers that can benefit from computational science can realistically do this today?
  - how does grid /distributed computing fit with large scale HPC applications?



# Where do researchers go for services?

- PI owned and operated cluster
- Campus Condo Computing
- Campus Research Computing
- Campus Condor Pool
- Regional Grids (eg NWICG)
- Open Science Grid: DOE/NSF, opportunistic access
- TeraGrid: NSF, competitively awarded allocations
- DOE ASCR: INCITE awards
- Commercial Cloud service providers
- There must be more ... what have I missed? where is NIH?



# Where do researchers go for services?

Answer: wherever they can get it, with the least amount of pain

- PI owned and operated cluster
- Campus Condo Computing
- Campus Research Computing
- Campus Condor Pool
- Regional Grids (eg NWICG)

### How many different:

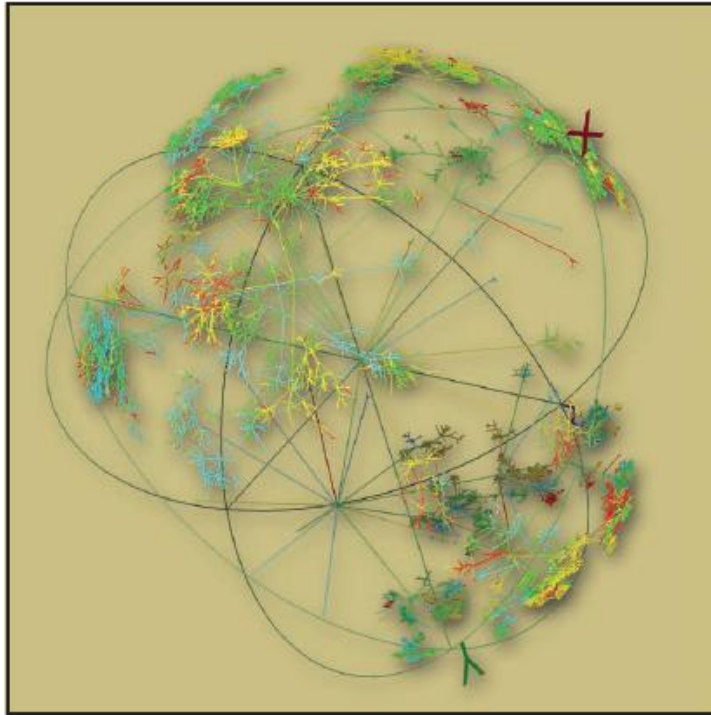
- service interfaces
- software stacks
- policy frameworks
- identities per researcher

...

where is The National Cyberinfrastructure?

- Open Science Grid: DOE/NSF, opportunistic access
- TeraGrid: NSF, competitively awarded allocations
- DOE ASCR: INCITE awards
- Commercial Cloud service providers
- There must be more ... what have I missed? where is NIH?

## CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY



National Science Foundation  
Cyberinfrastructure Council  
March 2007

and  
structure  
e and  
ision fo-  
mission  
tered,  
cipa-  
e, and  
es are  
cognize  
inclusive  
in sci-  
ps and  
ers in all  
d rely  
sment,  
is chap-  
goals  
e initia-  
ategy to

An effective computing environment designed to meet the computational needs of a range of science and engineering applications will include a variety of computing systems with complementary performance capabilities. NSF will invest in leadership class environments in the 0.5-10 petascale performance range. Strong partnerships involving other federal agencies, universities, industry and state government are also critical to success. NSF will also promote resource sharing between and among academic institutions to optimize the accessibility and use of HPC assets deployed and supported at the campus level. Supporting software services include the provision of intelligent development and problem-solving environments and tools. These tools are designed to provide improvements in ease of use, reusability of modules, and portable performance.

<sup>1</sup> A picosecond is  $10^{-12}$  second

<sup>2</sup> A petascale is  $10^{15}$  operations per second with comparable storage and networking capacity

...e density for iron oxide (FeO) within the local density approximation, with  
...ors represent the spin density, showing the antiferromagnetic ordering.

8.50 x 11.00 in



# The Open Science Grid

**A framework** for large scale distributed resource sharing  
addressing the technology, policy, and social requirements of sharing

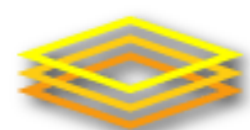
OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project. The project is funded by the NSF and DOE, and provides staff for managing various aspects of the OSG.

Brings petascale computing and storage resources into a uniform grid computing environment

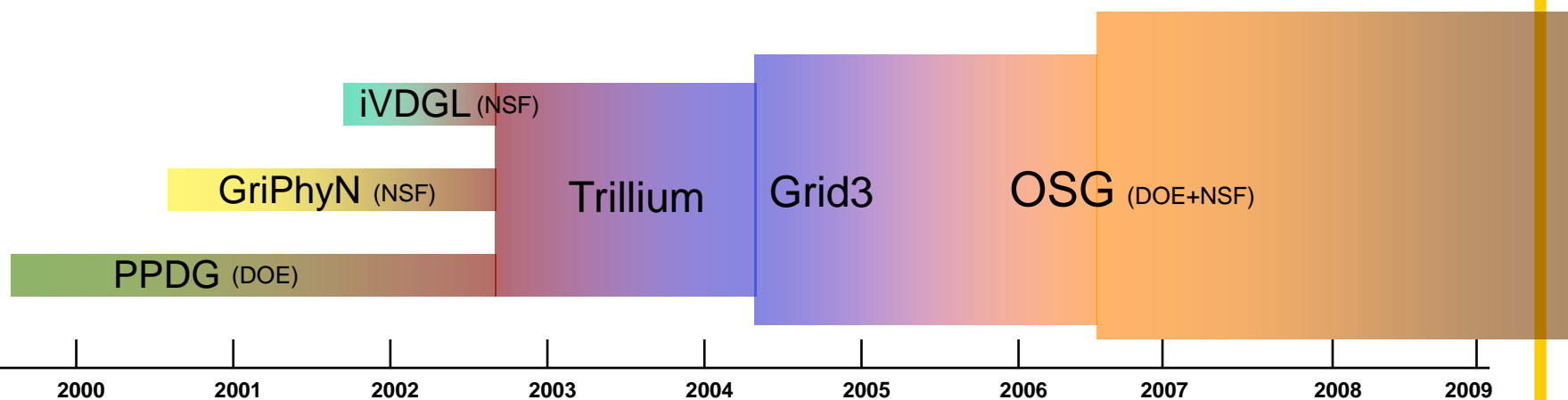
Integrates computing and storage resources from over 80 sites in the U.S. and beyond

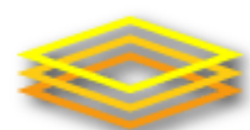






## Context: Evolution of Projects





## Using OSG Today

- Astrophysics
- Biochemistry
- Bioinformatics
- Earthquake Engineering
- Genetics
- Gravitational-wave physics
- Mathematics
- Nanotechnology
- Nuclear and particle physics
- Text mining
- And more...



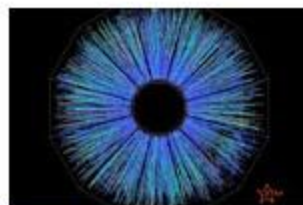
[ATLAS Detector](#)  
Copyright CERN  
[Permission Information](#)



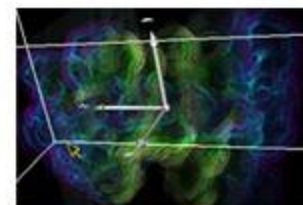
[SDSS Telescope](#)  
Image Credit Fermilab  
[Permission Information](#)



[CDMS photo](#)  
Image Credit Fermilab  
[Permission Information](#)



[STAR Collision](#)  
Image Credit Brookhaven  
National Laboratory/STAR  
Collaboration  
[Permission Information](#)



[BioMOCA Application in  
nanoHUB](#)  
Image Credit Shawn Rice,  
Purdue University  
[Permission Information](#)



[CMS Detector](#)  
Copyright CERN  
[Permission Information](#)



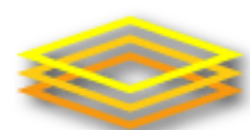
[Auger photo](#)  
Image Credit Pierre Auger  
Observatory  
[Permission Information](#)



[MiniBooNE photo](#)  
Image Credit Fermilab  
[Permission Information](#)



[DZero Detector](#)  
Image Credit Fermilab  
[Permission Information](#)

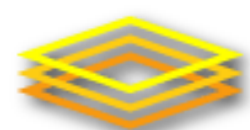


## Why should my University facilitate (or drive) resource sharing?

- Enables new modalities of collaboration
- Enables new levels of scale
- Democratizes large scale computing
- Sharing locally leads to sharing globally
- Better overall resource utilization
- Funding agencies care about this

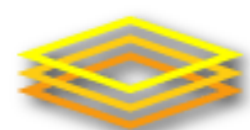
*At the heart of the cyberinfrastructure vision is the development of a cultural community that supports peer-to-peer collaboration and new modes of education based upon broad and open access to leadership computing; data and information resources; online instruments and observatories; and visualization and collaboration services.*

Arden Bement  
CI Vision for 21<sup>st</sup> Century introduction

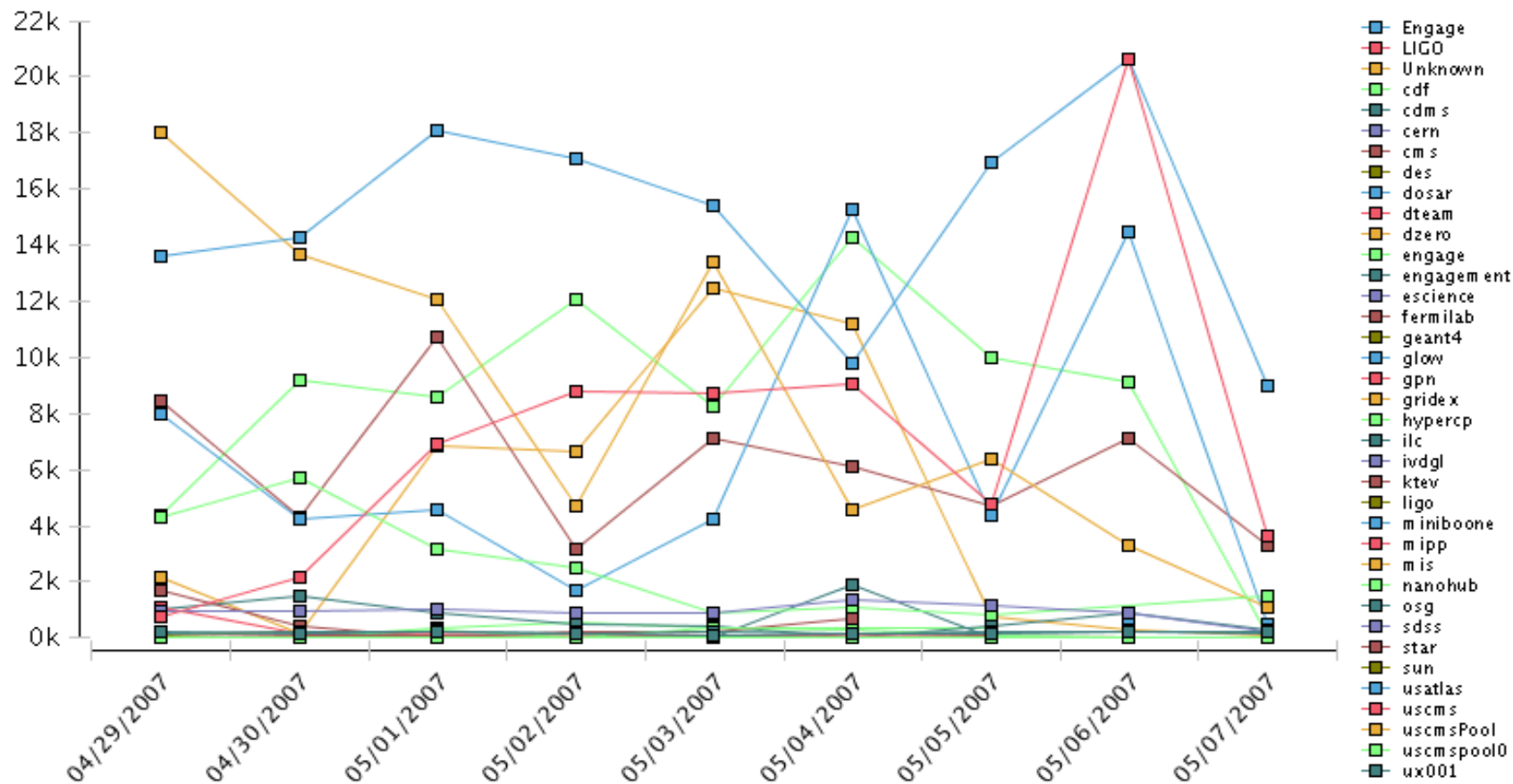


## OSG Engagement Mission

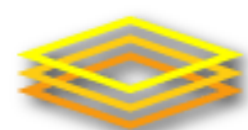
1. Help new user communities from diverse scientific domains adapt their computational systems to leverage OSG
2. Facilitate University Campus CI deployment, and interconnect it with the national community
3. Provide feedback and new requirements to the infrastructure providers



CPU Usage (process hours) By VO  
GratiaUser

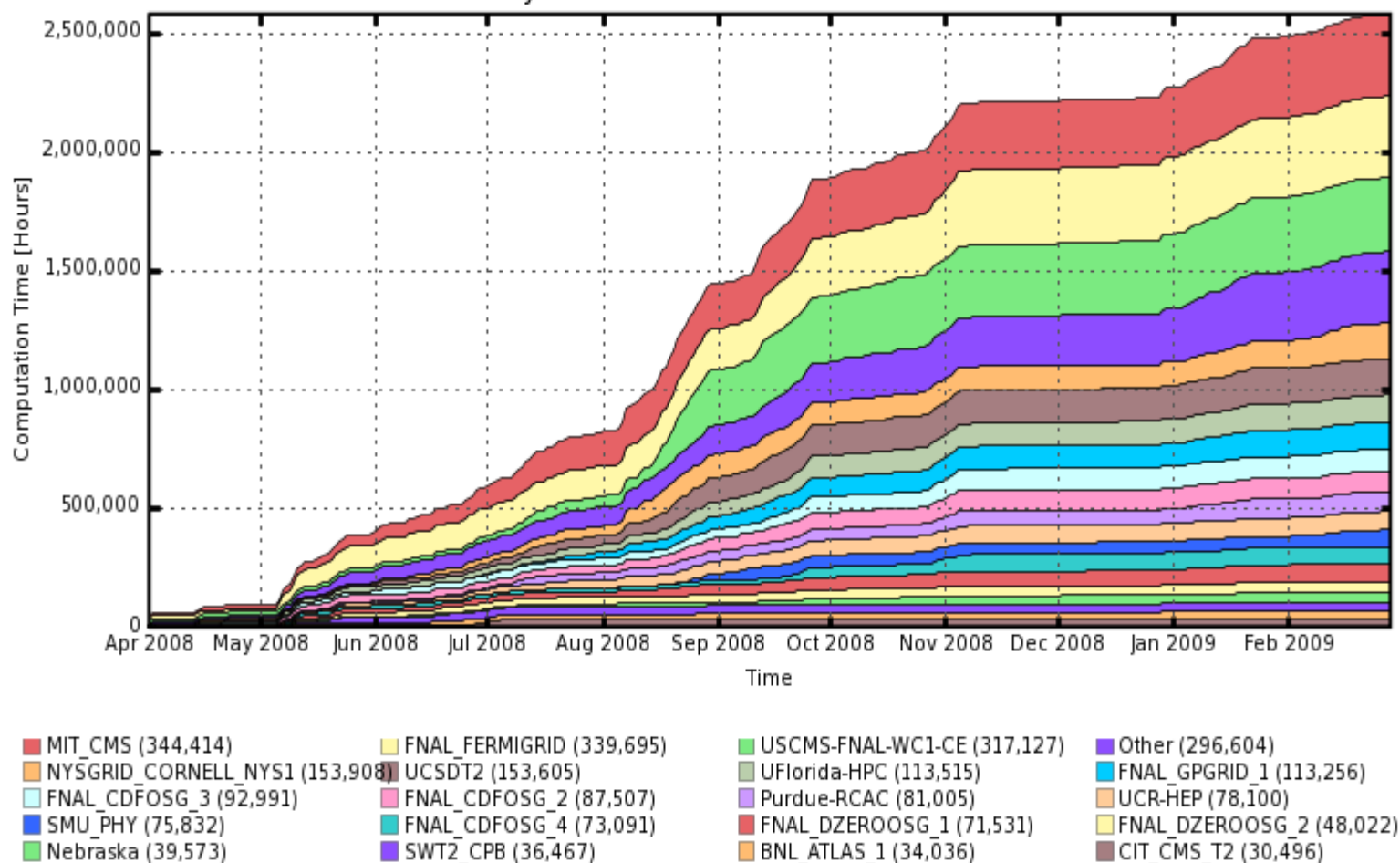


Date range: 2007-04-29 00:00:00 GMT - 2007-05-07 23:59:59 GMT



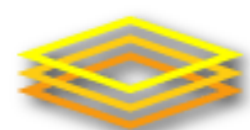
## Engaged Users Activity

Cumulative Hours Spent on Jobs By Facility  
332 Days from Week 13 of 2008 to Week 08 of 2009



Total: 2,580,785 Hours, Average Rate: 0.09 Hours/s





# Questions?

## OSG Engagement VO

<https://twiki.grid.iu.edu/twiki/bin/view/Engagement/WebHome>

**[engage-team@opensciencegrid.org](mailto:engage-team@opensciencegrid.org)**



**E N D**