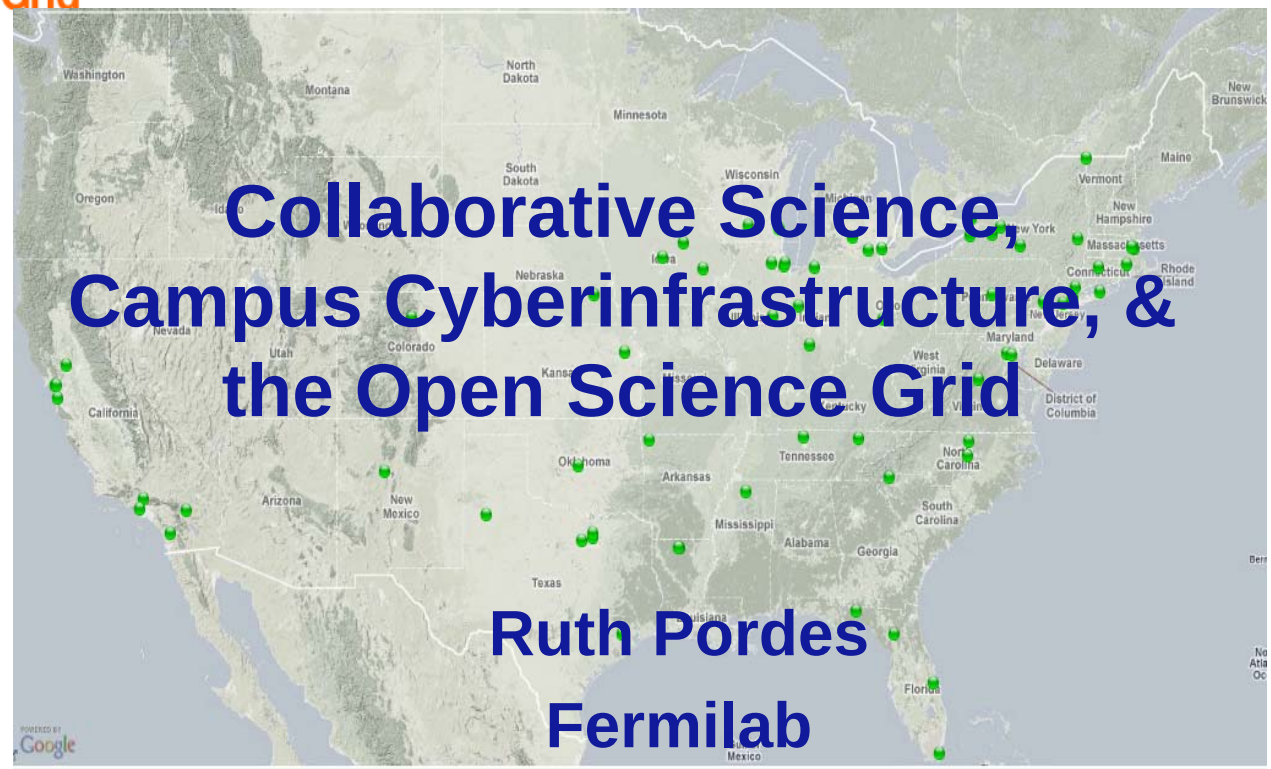
Open Science Grid



Collaborative Science, Campus Cyberinfrastructure, & the Open Science Grid

Ruth Pordes

Fermilab

Open Science Grid

Outline

- OSCER, Fermilab, OSG
- Sharing of Software,Computing, Knowledge
- Futures

Open Science Grid

Fermilab

Open Science DOE Lab that supports world wide scientific collaborations of 1000s of physicists and a range of astrophysics, HEP, theory and accelerator science.

Accelerators Detectors Data Analysis

Energy Frontier Intensity Frontier Cosmic Frontier

Serves ~4,000 university users; internally ~200 in IT department: (Astro)Physicists, IT, Engineers, Computer & Computational Scientists. Participations in PhD programs in accelerator science; member of the DOE Computional Science Graduate Student Fellowship Program.

OSG

provides a cross-domain self-managed national ***distributed high-throughput computing facility***…

that brings together ***campus and community infrastructures at all scales…***

to provide services, support and facilitate the needs **of scientific and research communities at all scales…**

Does not own processing, storage, networking hardware.

Does not develop software.

Wide Area Science Distributed Computing Social Network

A Consortium with more than 100 institutions of Scientists, IT and Software Providers.

A 5 year funded project ($6M/yr) across NSF and DOE SciDAC-2 with staff at 16 institutions.

Staff provides all aspects of the core distributed facility: software, operations, user and administrator support, security, training, collaboration with peers.

Consortium members make significant contributions. Most active are Physics Collaborations: HEP, NP, LIGO, who are prepared to collaborate with and support other programs and disciplines.

Active partnerships with European projects, ESNET, Internet2, Condor, Globus…

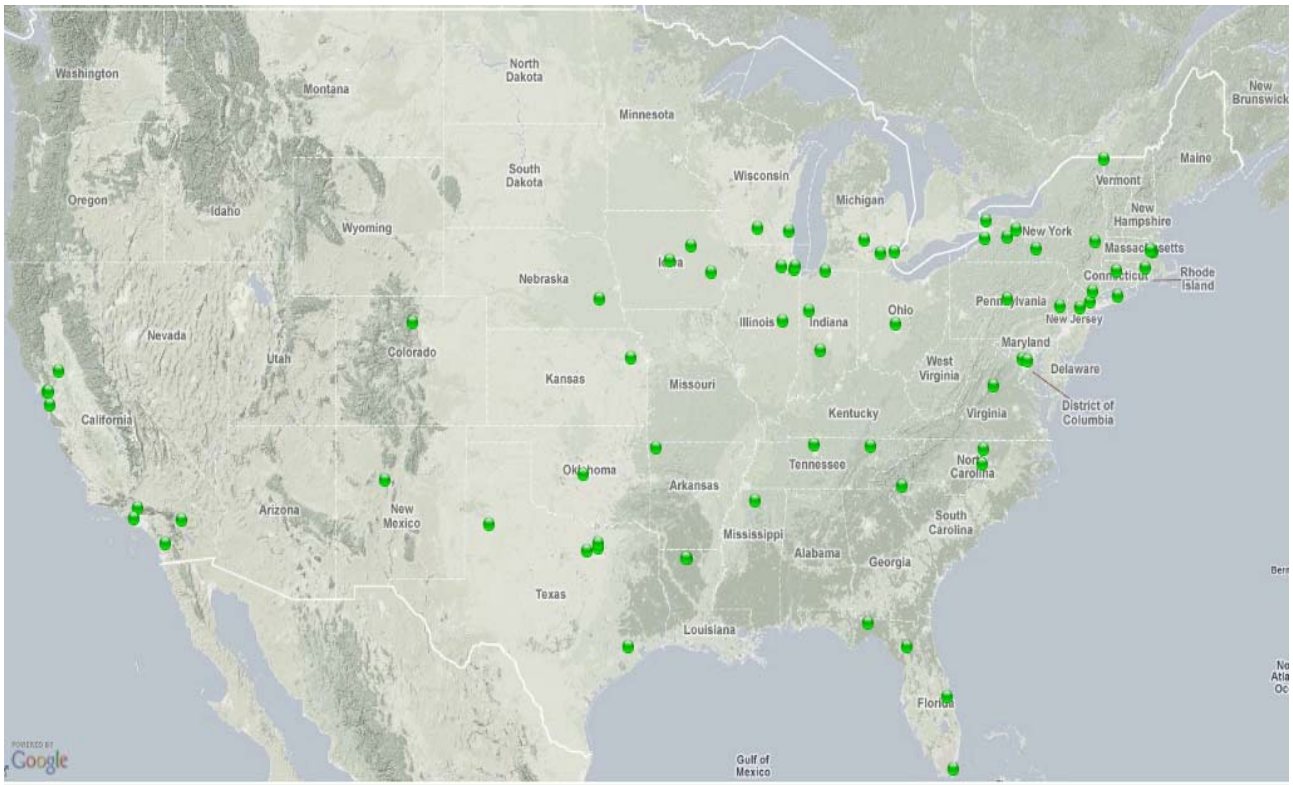
Virtual Organizations

OSG works with Virtual Organizations or Communities

There are 30 VOs in OSG spanning scientific, regional, campus, international and education,

There are specific “OSG owned VOs” to accommodate individual users

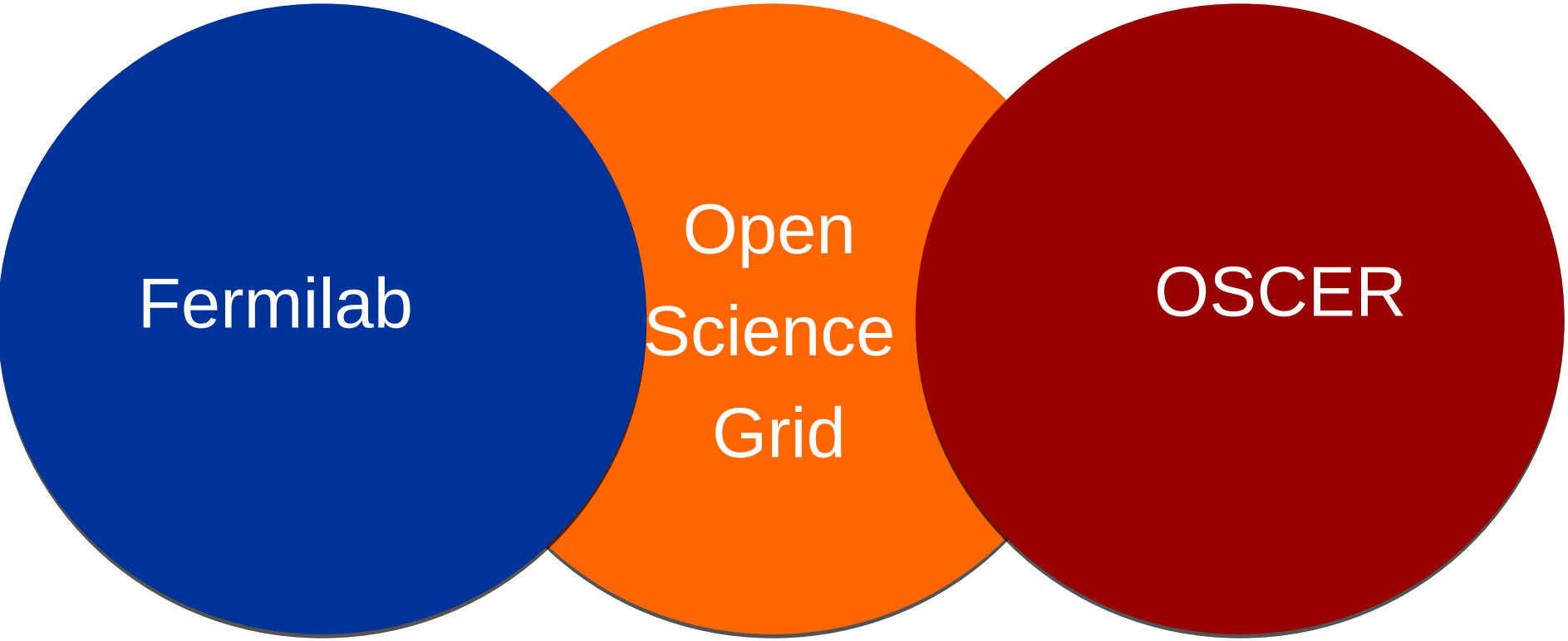
Map of Accessible Resources



Why User map not yet feasible.

Open Science Grid

The Distributed Computing Model



Federated Autonomous CyberInfrastructures

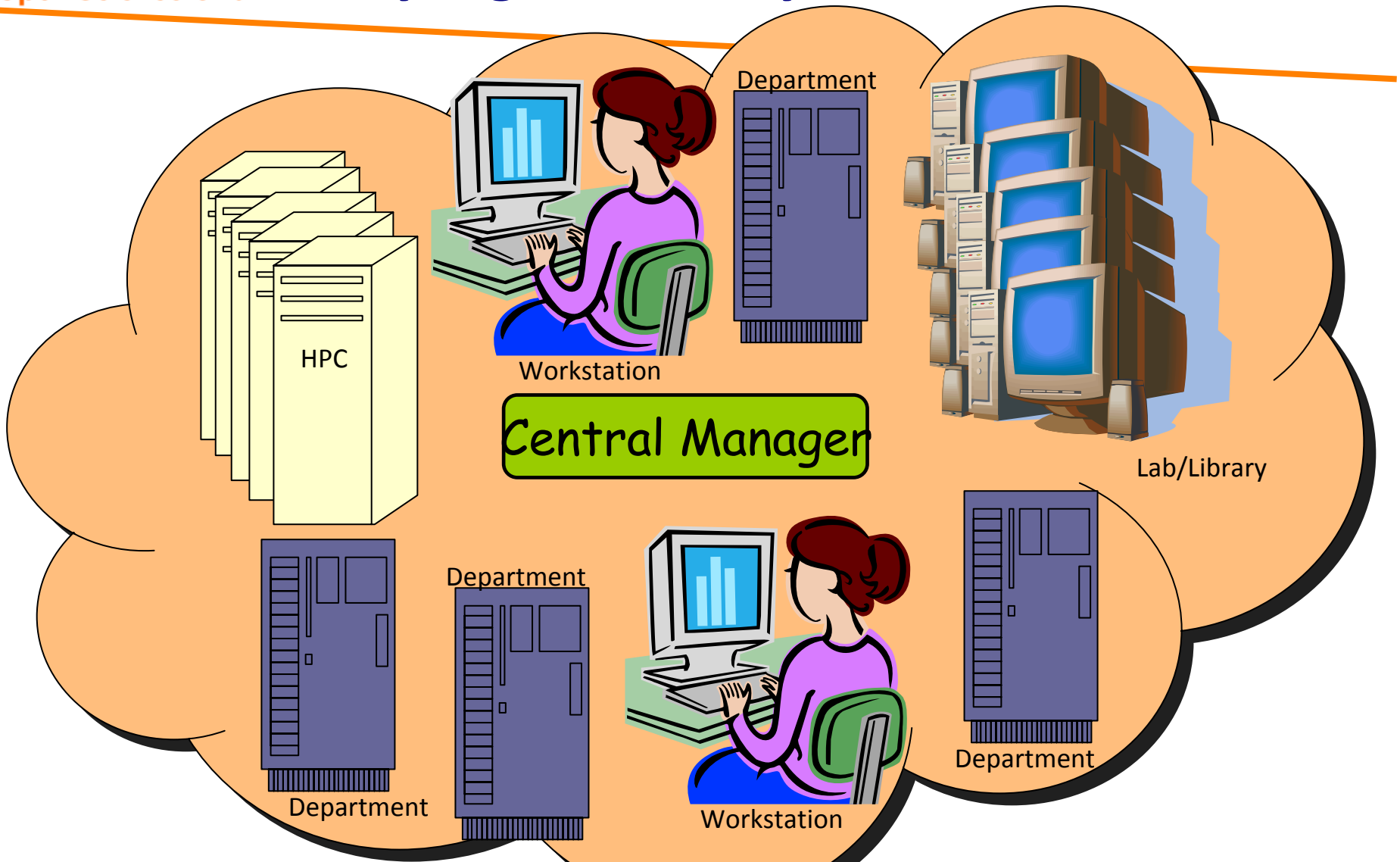
National
Cyber-
Infrastructure
Part of World-Wide
Infrastructure

Campus Grid &
Large Scale Science
Community Grids &
DOE Lab IT Facility

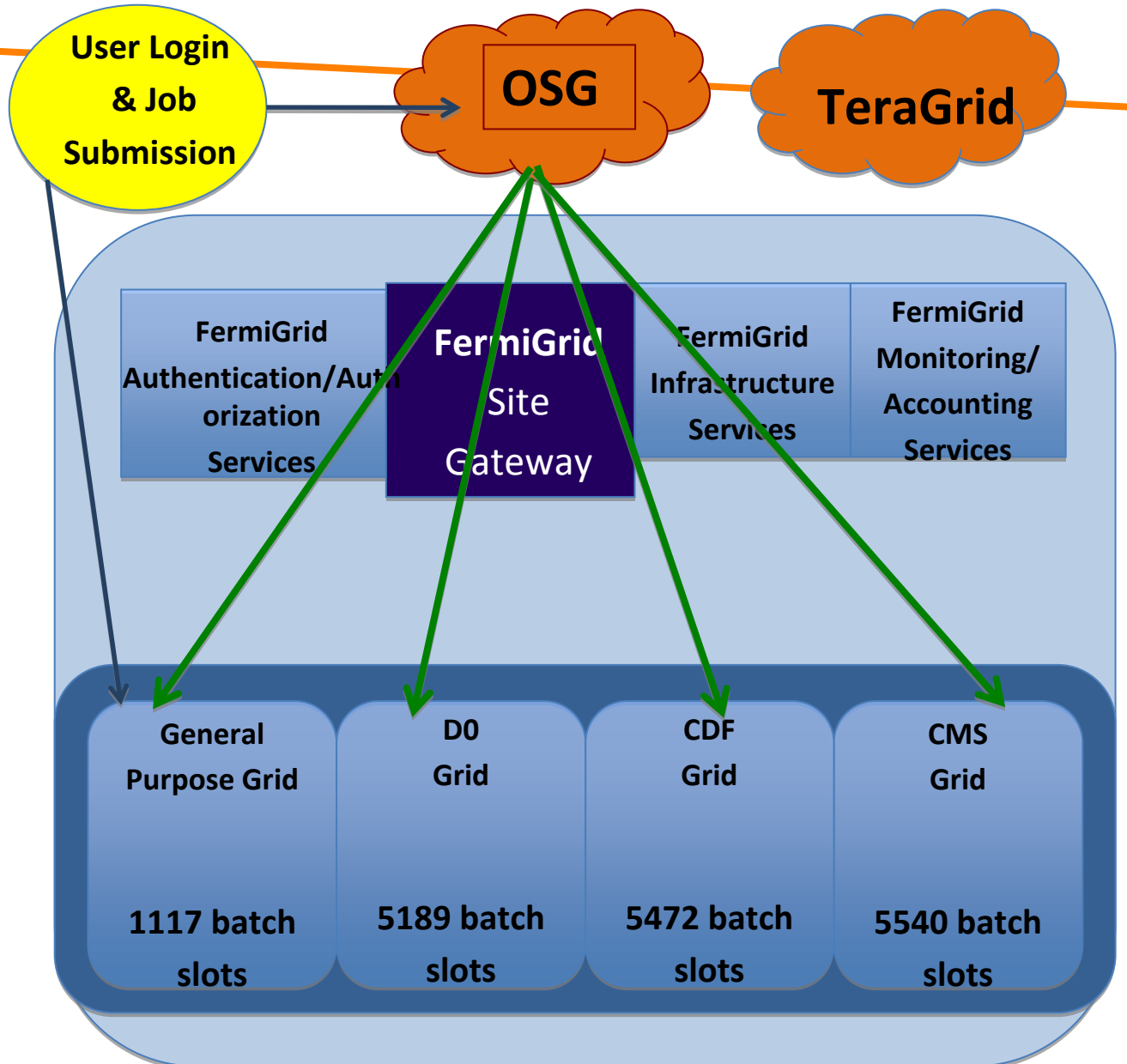
Faculty, Cross-Campus &
Regional Grids
Students, Educators,
Researchers, CS…

Several Current Models for Campus Grids (in OSG)

A Campus Condor pool with no OSG connectivity e.g University of North Carolina,

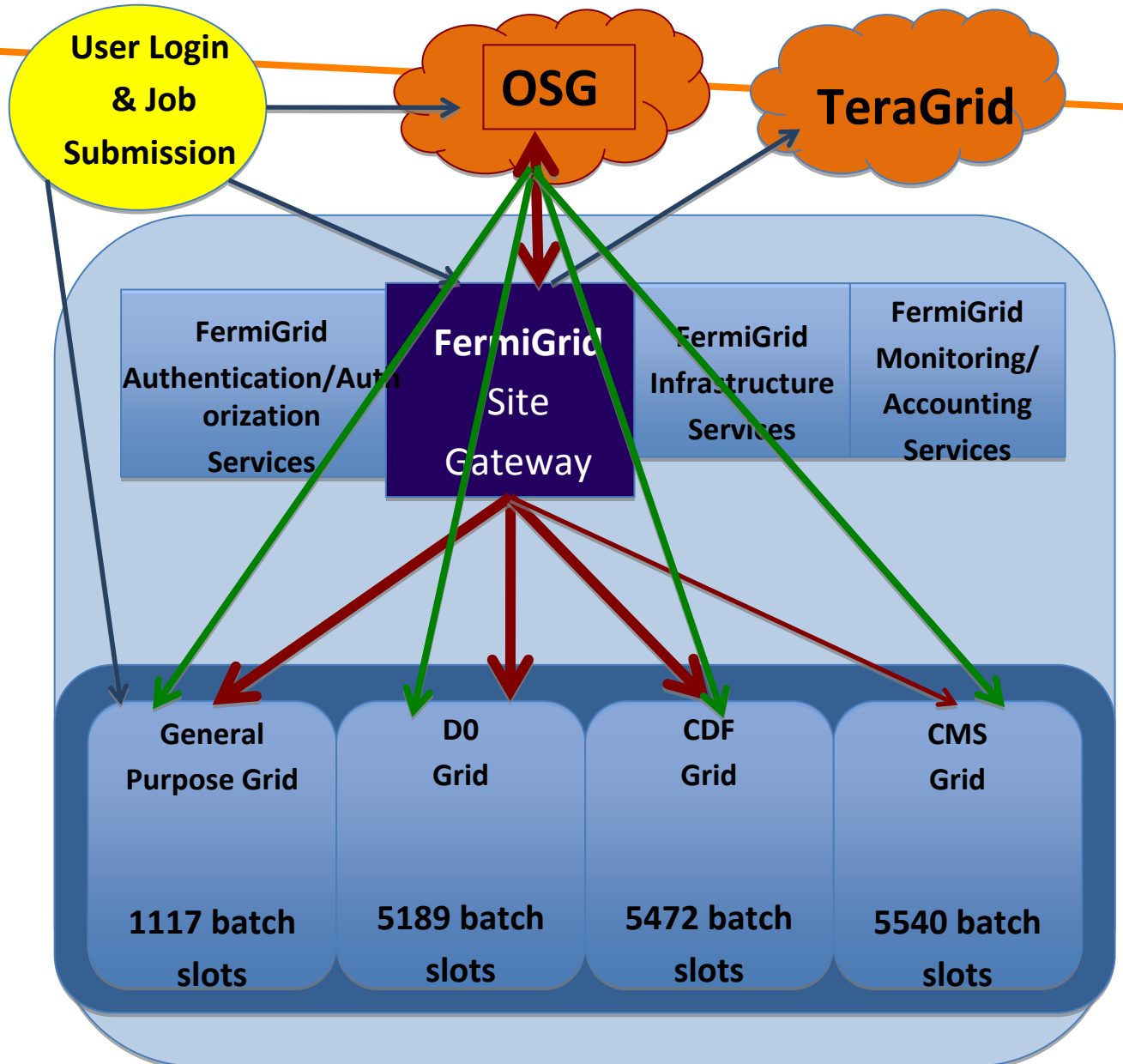


FermiGrid : Cross Campus sharing with Each Cluster a Node on the Wide Area Grid



Total batch slots: 17,318

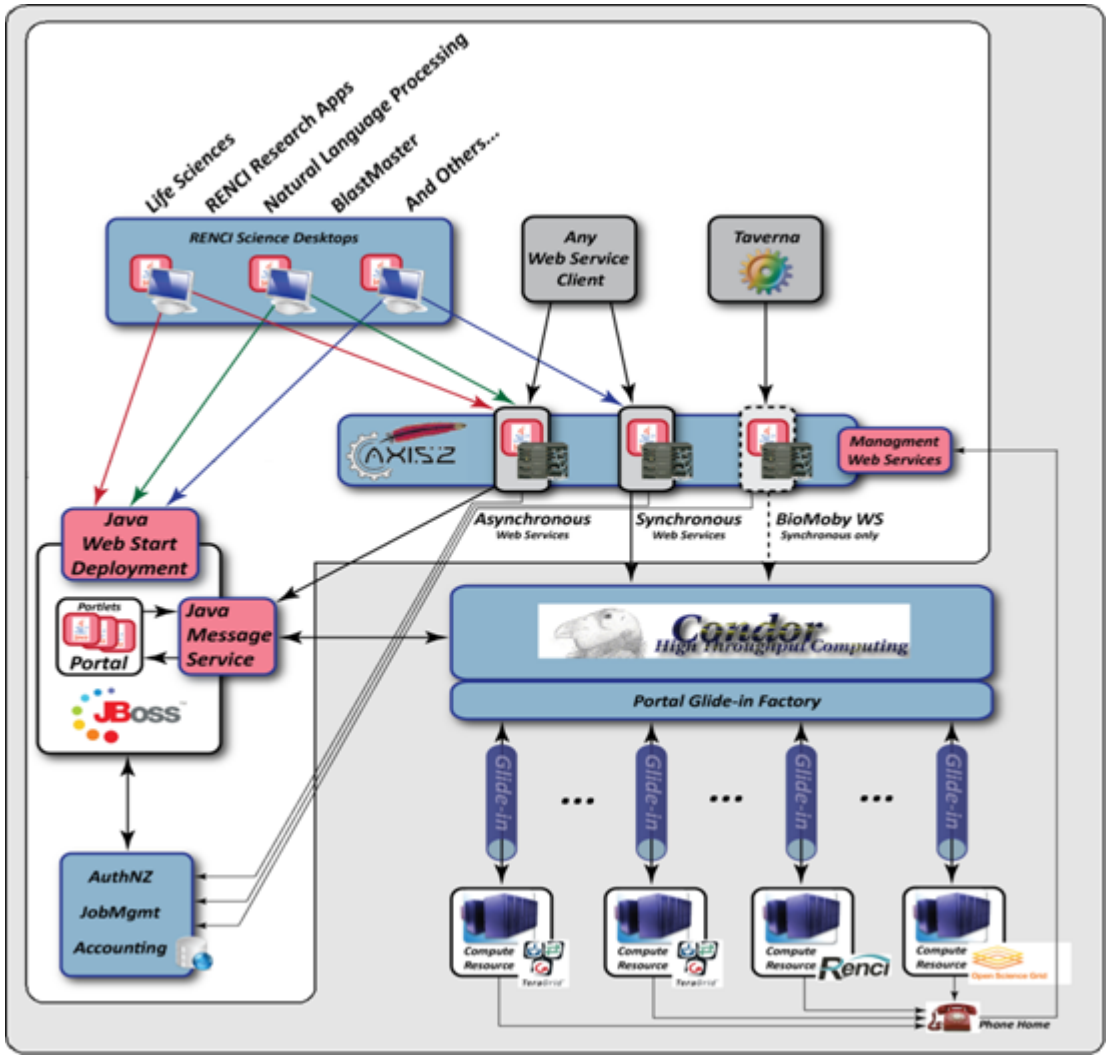
Cross Campus sharing with Each Cluster also a Node on the Wide Area Grid



Total batch slots: 17,318

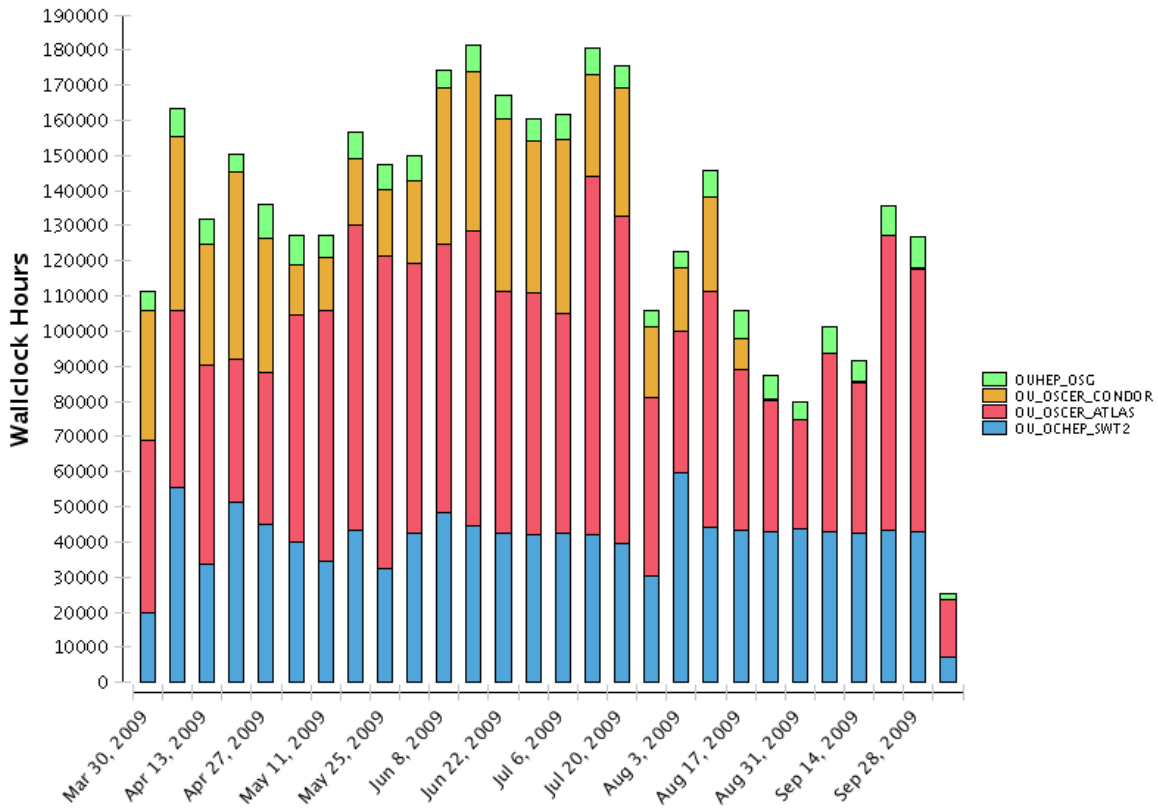
Open Science Grid

University of North Carolina



OSCER

- A “mix” several clusters as one Condor Pool, with several “gateways” to the Wide Area Grid.
- Averaging ~100 CPUhours/hour for OSG use – sharing unused local compute cycles



Open Science Grid

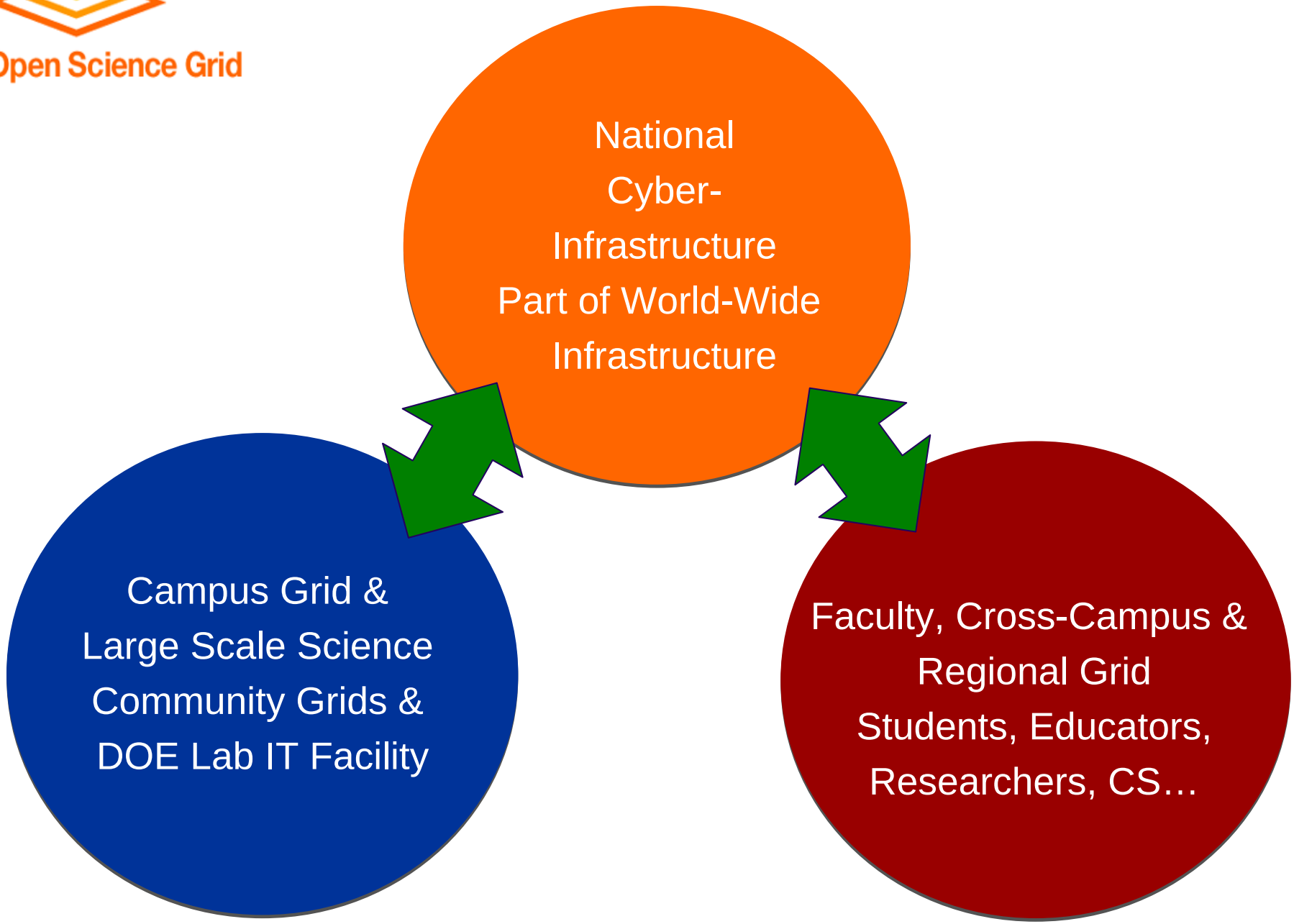
OSG Helps

University researchers, faculty and students to **interact with "national CI"** including TeraGrid, OSG, other Regional grids etc.

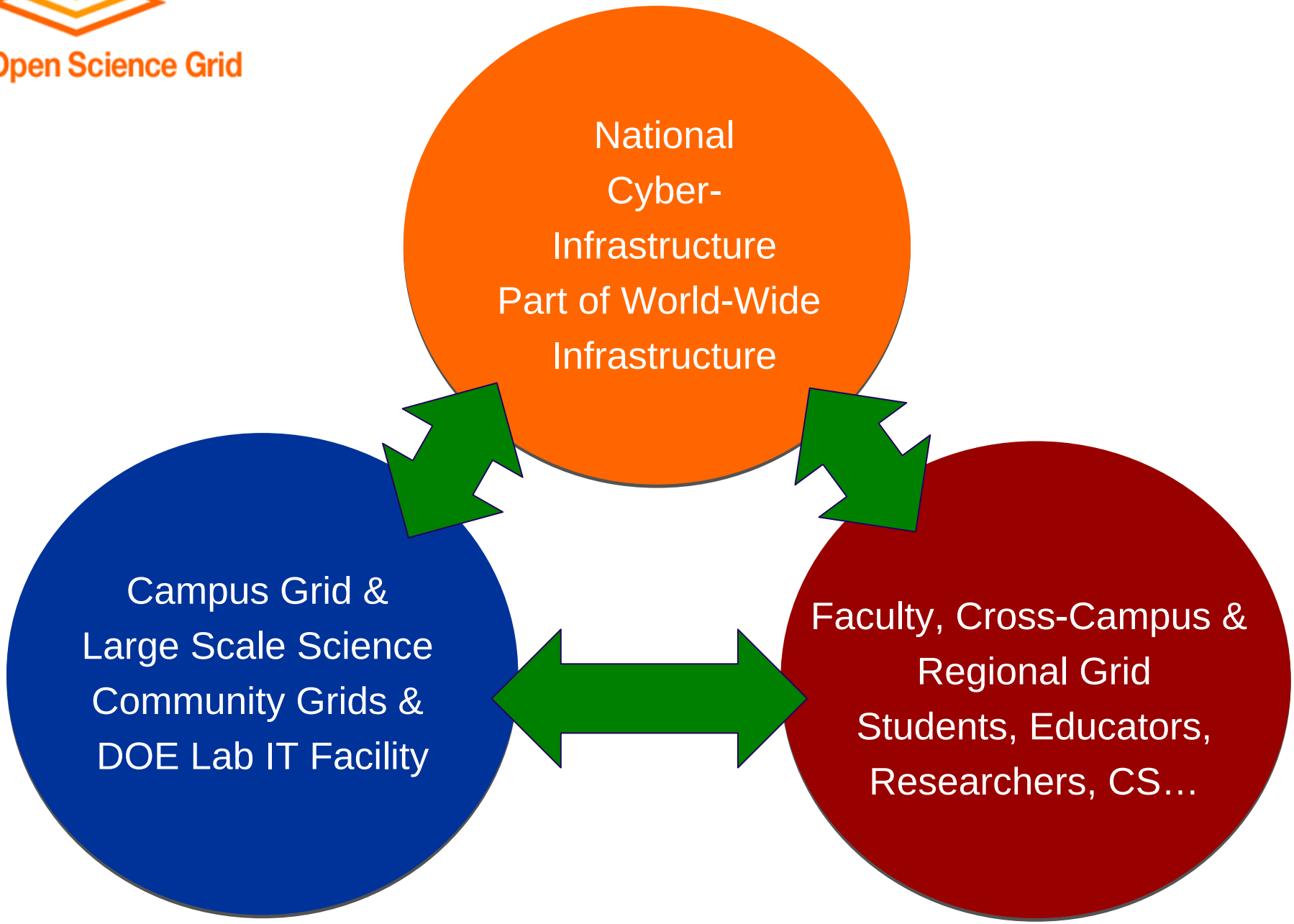
Organize the campuses to change their culture and **Share Resources locally.**

Enable campuses and users to send jobs and data "on-demand" between the local and remote facilities, becoming **full members of the nation-wide shared cyber-infrastructure.**

Open Science Grid



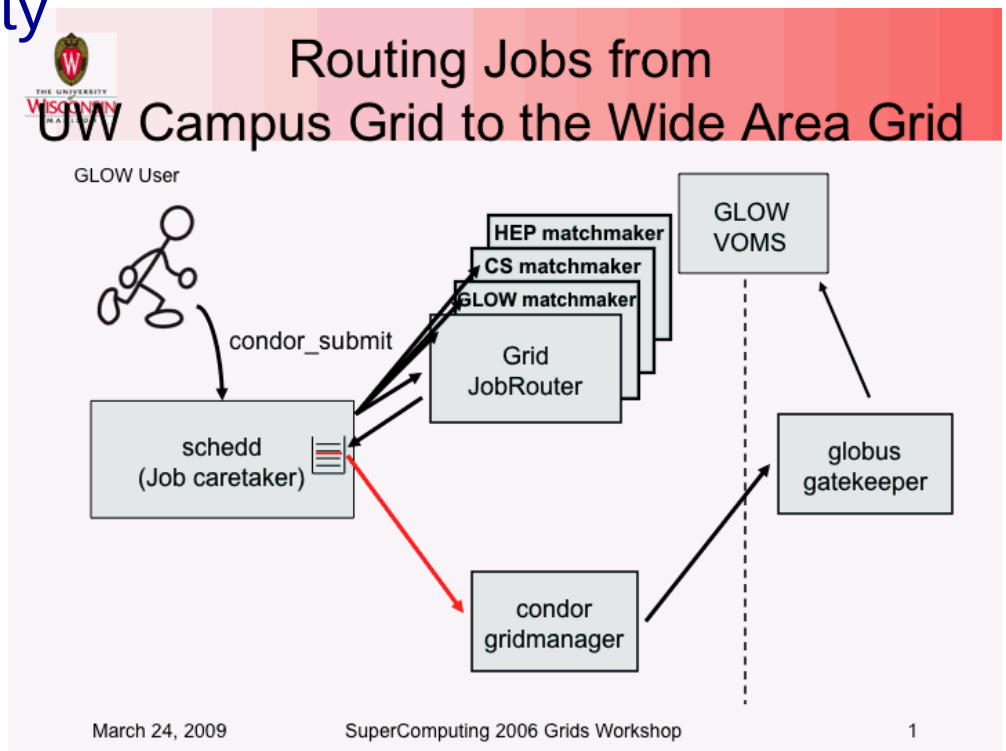
Open Science Grid



Condor job router – Grid Laboratory of Wisconsin

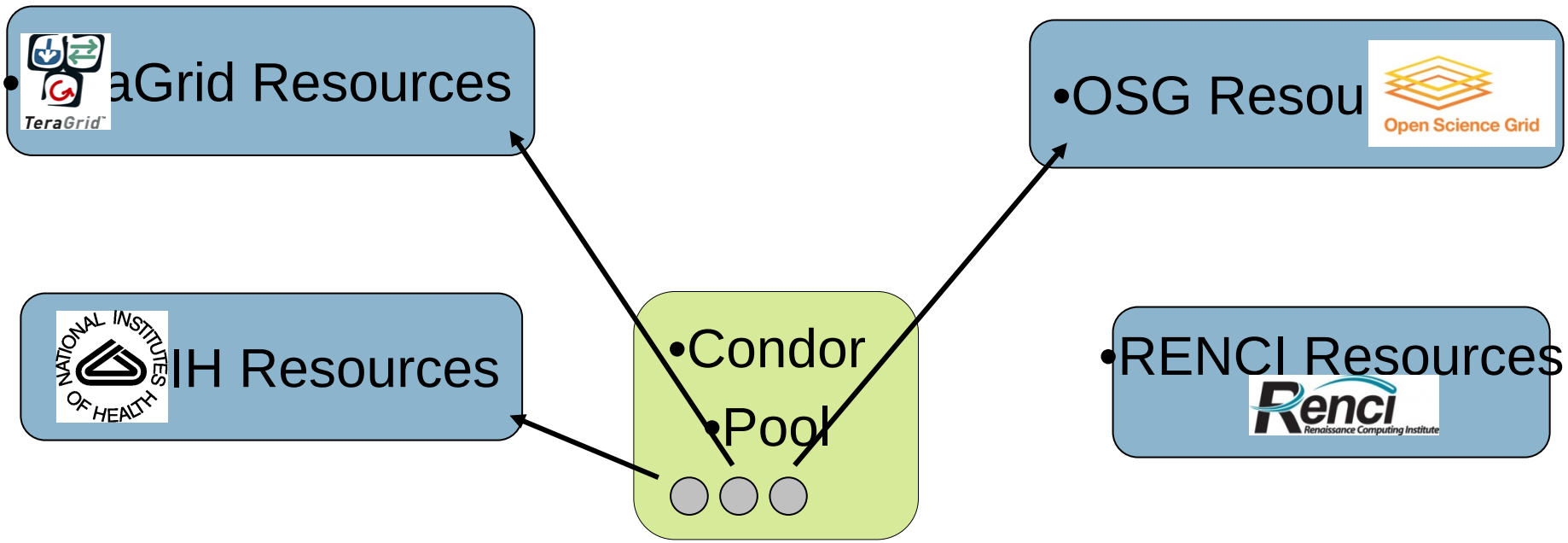
Automated way to let jobs run on a wider array of resources

- Transform jobs into different forms
- Reroute jobs to different destinations
- Handle security



Glide-ins at RENCI

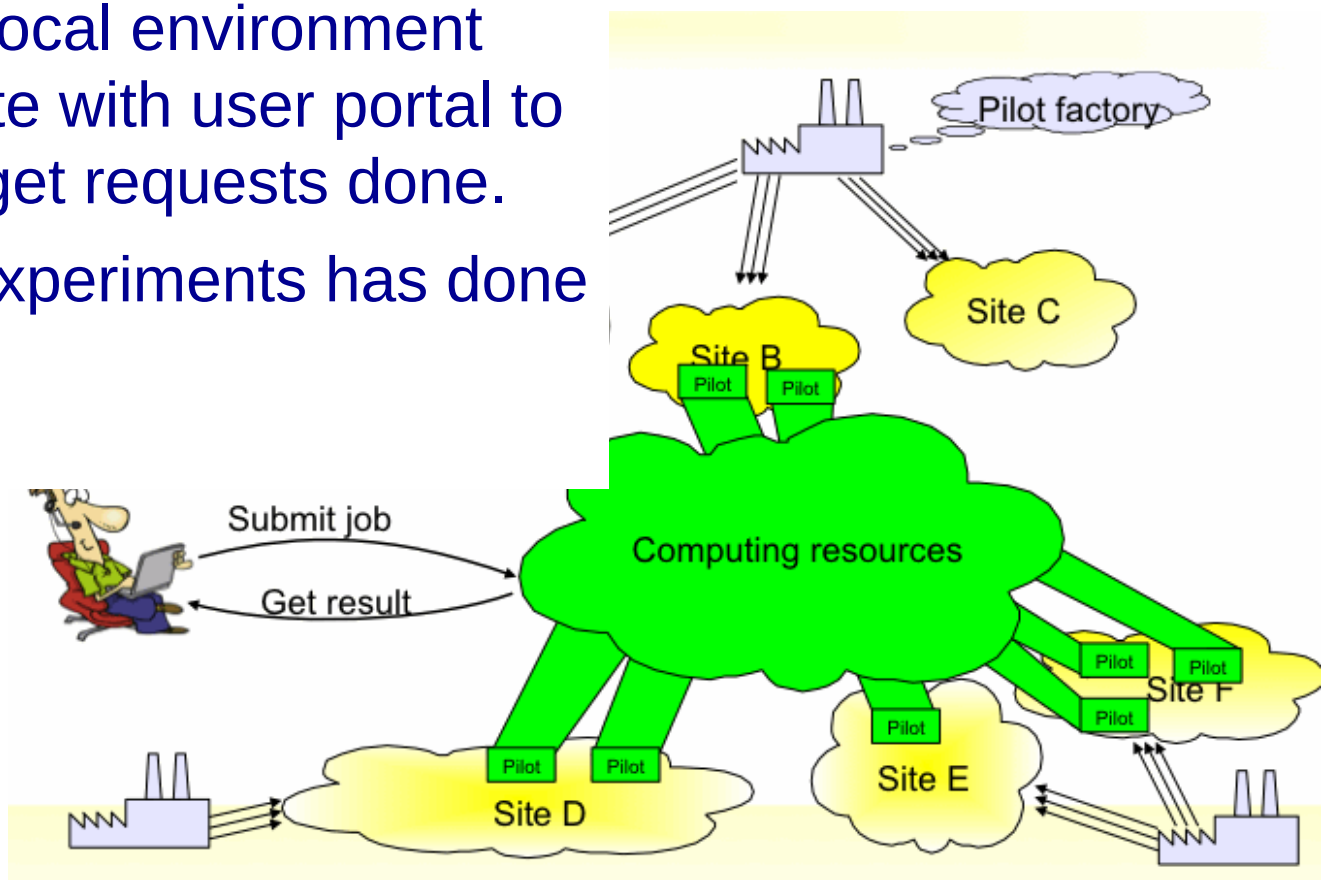
Temporarily join remote machines into local Condor pool

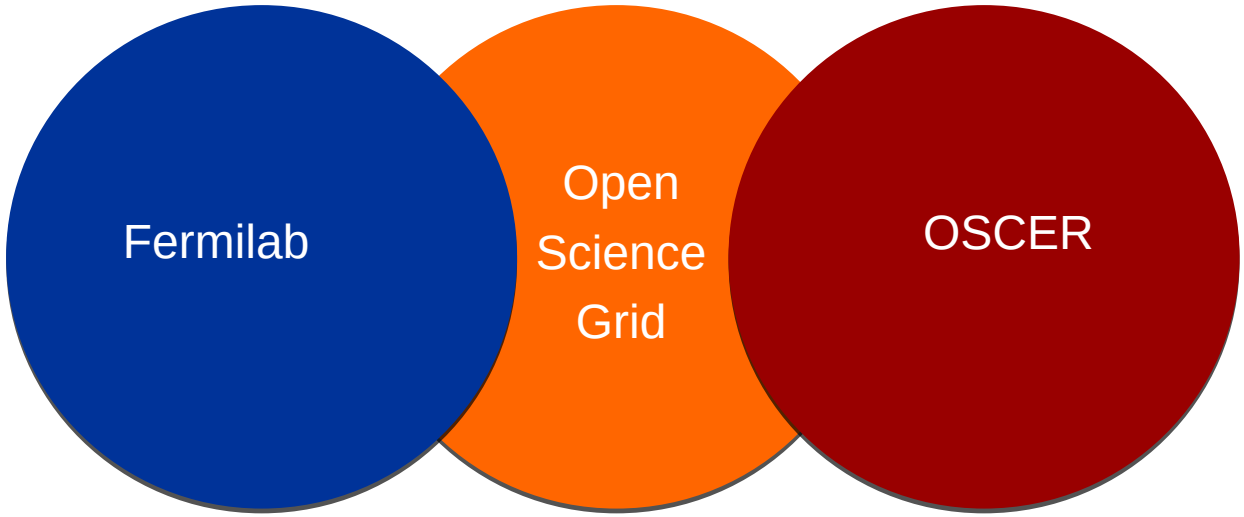


Overlaid Workspaces

Deploy “agents” on each site which learn about the local environment and communicate with user portal to most efficiently get requests done.

Each of the LHC Experiments has done this (differently)





Tevatron Physics

LHC Physics

Support for Science locally and broadly

Tevatron Physics

OU is a DZero Collaborator & important part of the Dzero Community SAMGrid, hosting data serving and catalog services.

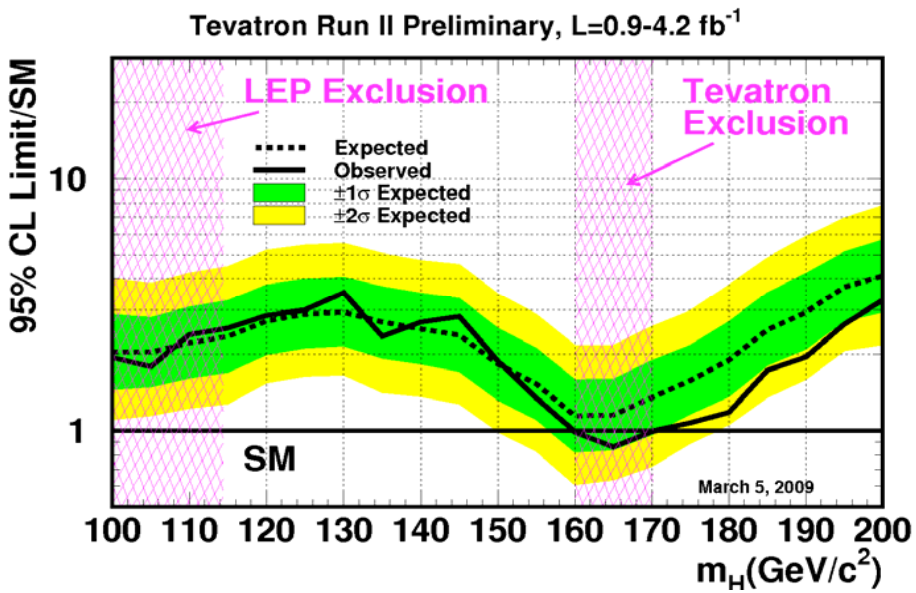
CDF and D0 Experiments profit from LHC resources through adapting to and using OSG. Interesting issue how to “quantify” the value gained.

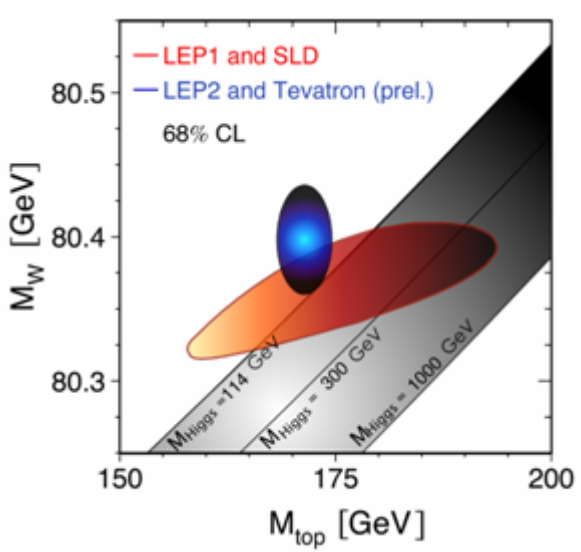
Brazilian institutions are important partners for both Fermilab and OSCER/DO.

RESULTS from
Supporting the Search for the Origin of Mass

Shrunk the Higgs mass window from a few hundred GeV to ~ 40 GeV. Thus providing a 135GeV +- 15% target region for future discoveries.

(press release March 2009).





- Direct Higgs Search at Tevatron
- now excluding high mass higgs.

- Top and W masses now known to
- 0.3% and 0.75% respectively.

=> Constrain Higgs via Quantum effects.

US LHC Physics

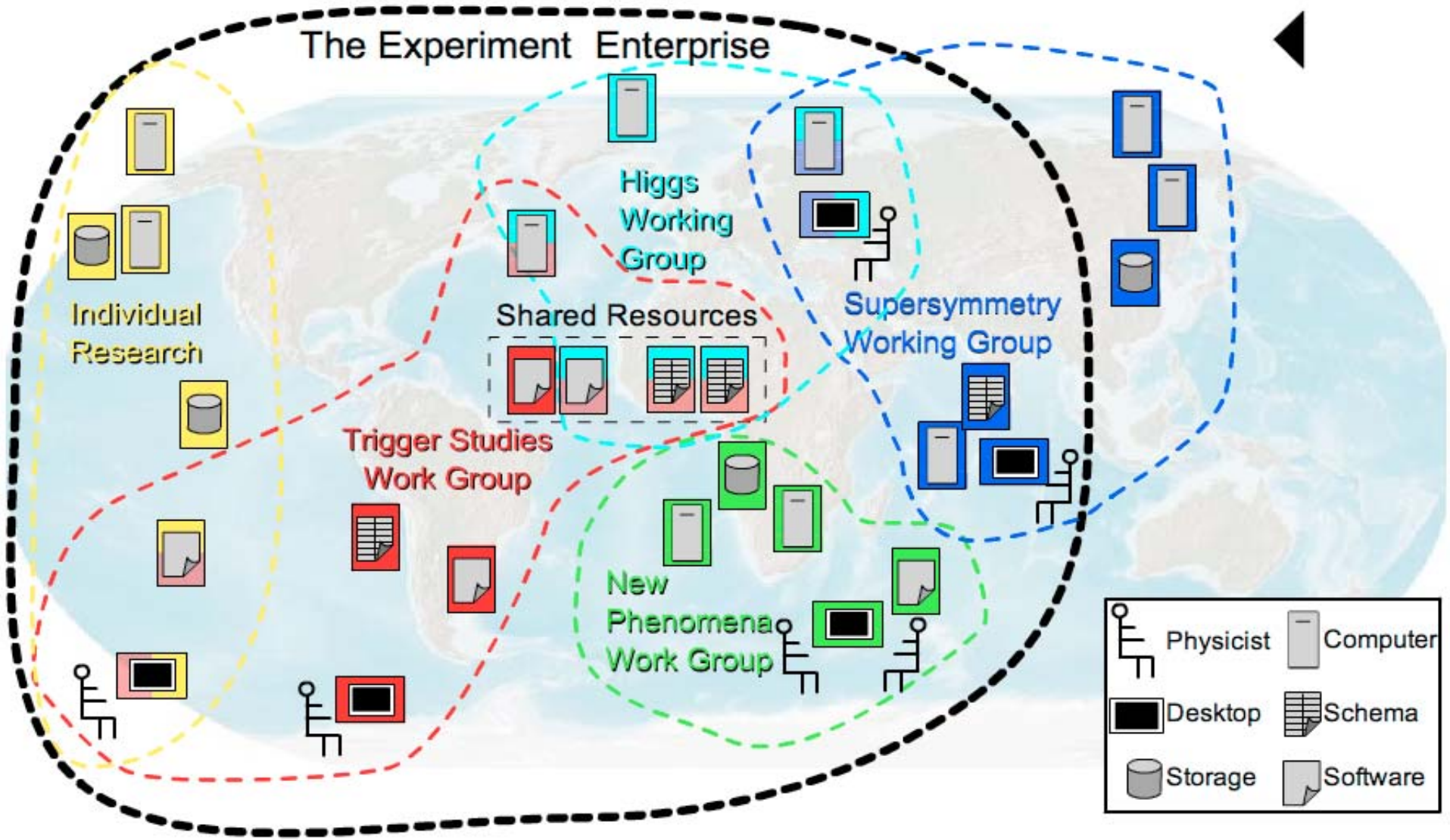
OU is US ATLAS Tier-2 site.

Fermilab is US CMS Tier-1.

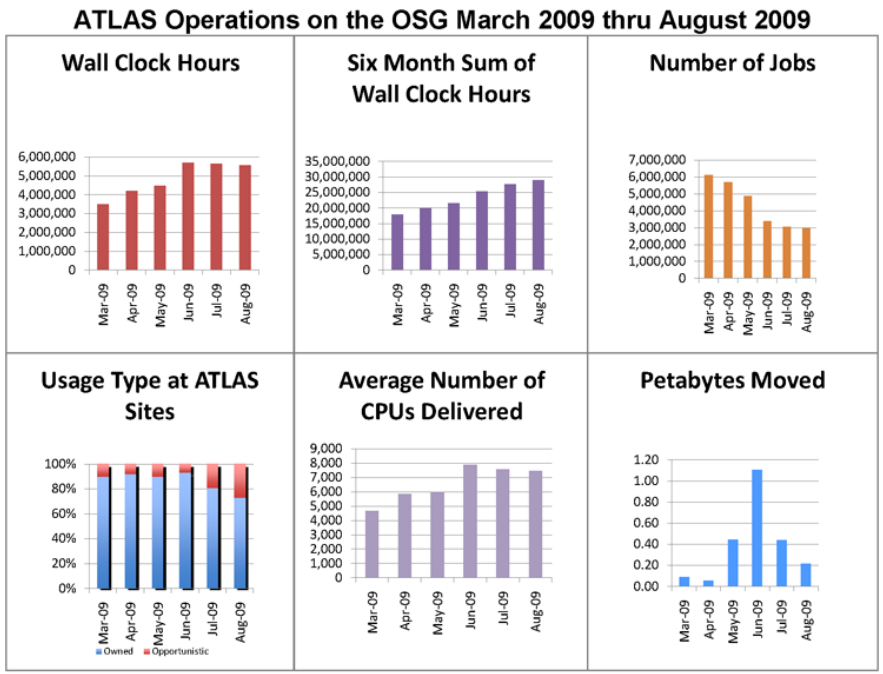
OSG Contributes to WorldWide LHC Computing Grid by supporting US ATLAS and US CMS.

South/Central American institutions important collaborators for experiments and OSG.

The LHC: WorldWide Community Overlays on Shared Facilities (circa 2005 & still true today)



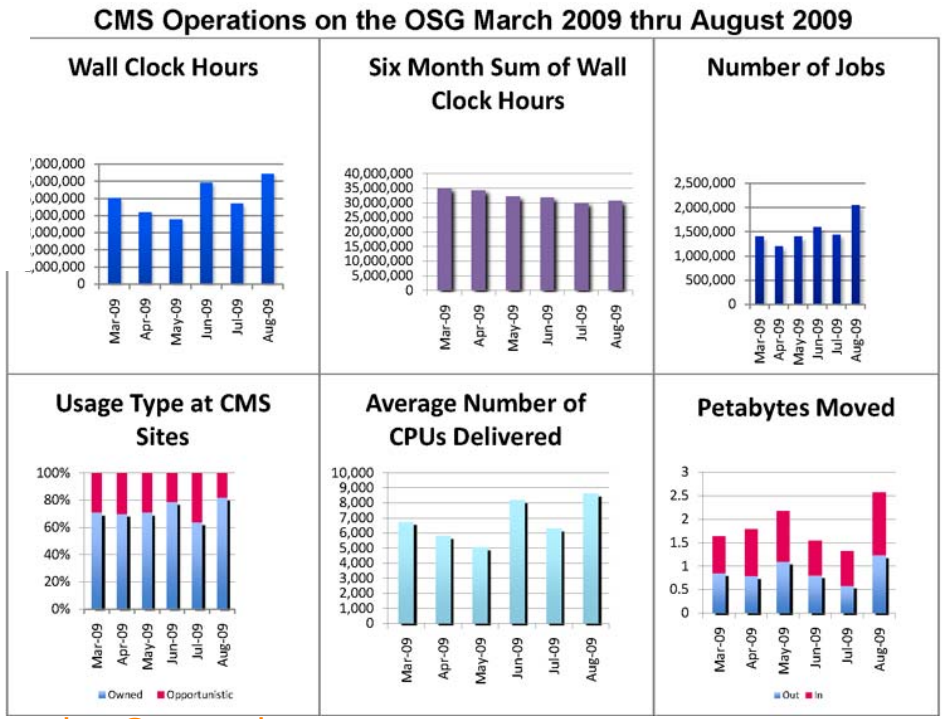
Scale of LHC Usage in US



Implies ~

300,000 jobs a day

3 TeraBytes moved each hour



Across 3 DOE labs,

15 large Universities,

10 smaller Universities

OSCER Contributes

User Communities (from 2008 talk)

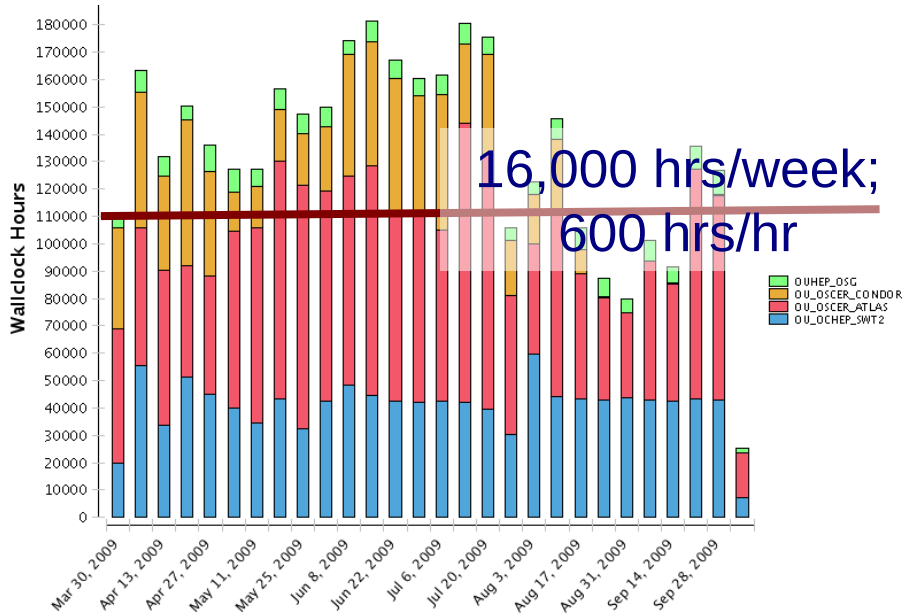
- Aerospace & Mechanical Engr
- NEW! Anthropology
- Biochemistry & Molecular Biology
- Biological Survey
- Botany & Microbiology
- Chemical, Biological & Materials Engr
- Chemistry & Biochemistry
- Civil Engr & Environmental Science
- Computer Science
- Economics
- Electrical & Computer Engr
- Finance
- Health & Sport Sciences
- History of Science
- Industrial Engr
- Geography
- Geology & Geophysics
- Library & Information Studies
- Mathematics
- Meteorology
- Petroleum & Geological Engr
- Physics & Astronomy
- NEW! Psychology
- Radiological Sciences
- Surgery
- Zoology

More than 150 faculty & staff in 26 depts in Colleges of Arts & Sciences, Atmospheric & Geographic Sciences, Business, Earth & Energy, Engineering, and Medicine – with more to come!

OSCER State of the Center Address
Tuesday October 7 2008

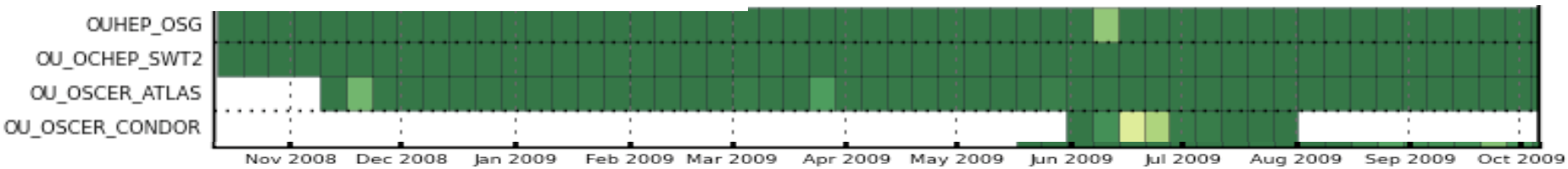
7

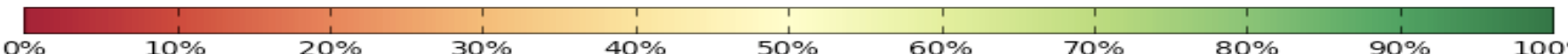
Computing Cycles



16,000 hrs/week;
600 hrs/hr

Sustained Reliability

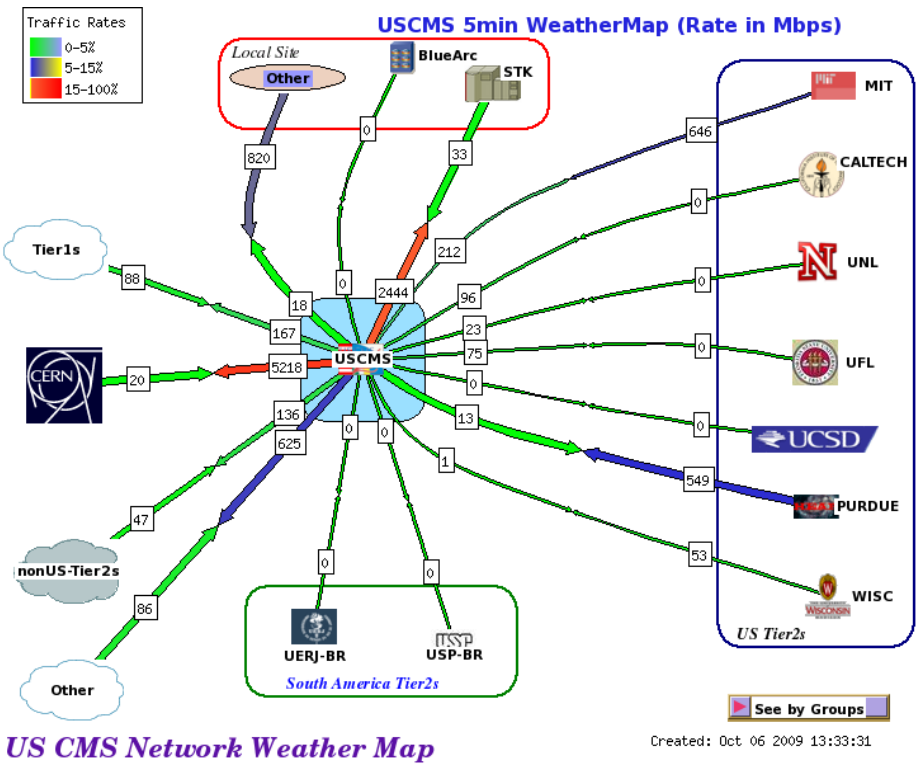




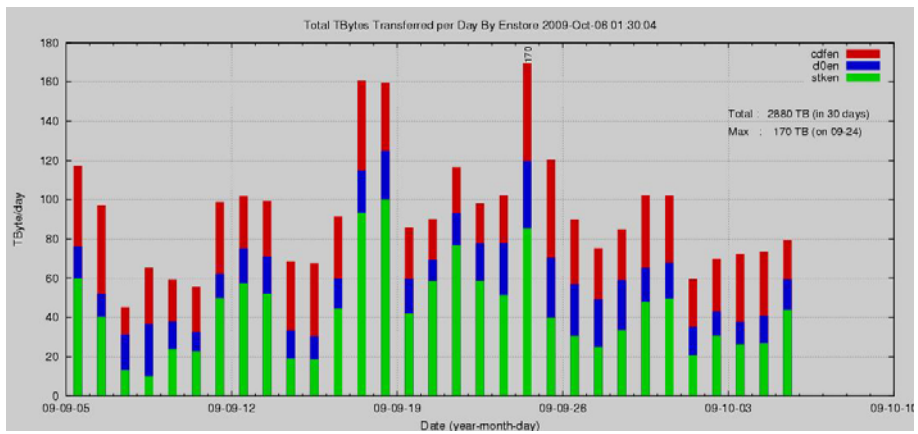
Fermilab brings Large Data Storage, Network and Data Movement expertise

Total User Data on Tape (Cdfen, D0en, Stken) :	20976.768 TB

Data Input and Output Hub



Large Daily I/O to/from Tape



Acting as an Agency

..supports and brokers relationships/expectations between user communities & resources, services, software.

Some other users

Gravitational Wave: Einstein@Home, gradually progressing #2 in the world

Protein structure prediction: Toyota Institute.

Weather Research Forecasting: U. North Carolina, University of Nebraska.

Structural Biology Predictions: Harvard Medical School.

Nanotechnology Simulation and Modelling: Purdue University and Nanohub collaboration.

Molecular Dynamics: U of Buffalo, Argentina

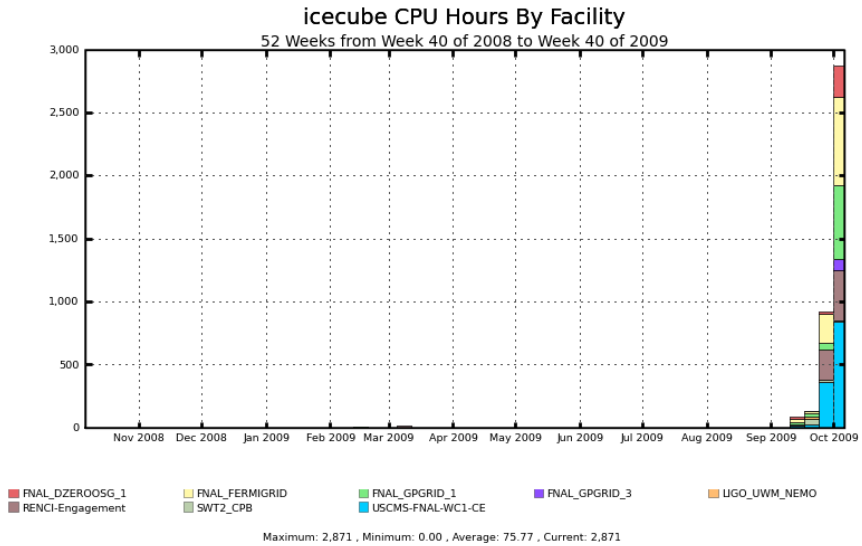
Theoretical Nuclear Physics: Duke University..

Text Mining: U. North Carolina

Exploring Mathematical algorithms: University of Colorado

Once applications are adapted, communities ramp up to use multiple sites quickly

IceCube, a telescope under construction at the South Pole, will search for neutrinos from the most violent astrophysical sources: events like exploding stars, gamma ray bursts, and cataclysmic phenomena involving black holes and neutron stars.



Sharing of Software and Knowledge

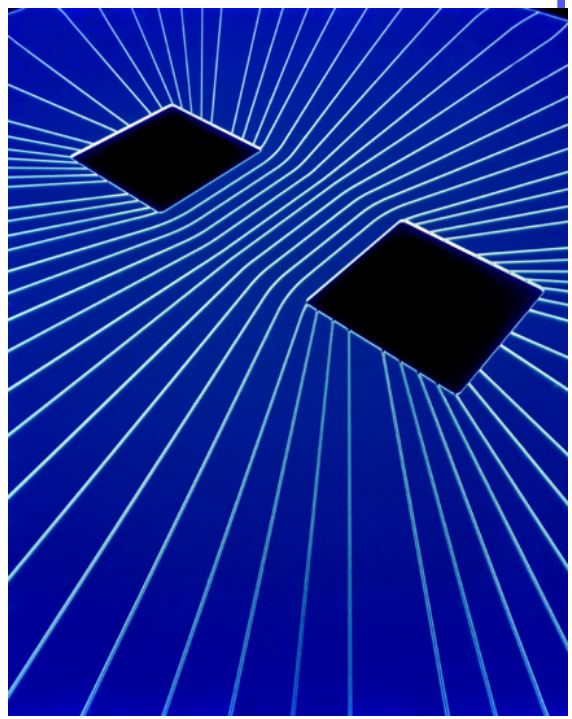
Computing, data, storage rely on software to be useful.

OSG provides a common software set which is packaged, tested, distributed for many different OS and provided in different configurations for Users, Communities, Processing, Storage, and Data servers.

Software built and tested on ~15 Linux variants.

System testing done on separate test Grid with more than 15 sites and 10 science communities participating

Open Science Grid



Alain Roy,

The software coordinator. He is a computer scientist in the Condor Project.

(He is a master bread maker)

Architectural Layers – whether Local or a Campus or Wide Area

Open Science Grid

e.g. OSG

Users, User Applications,

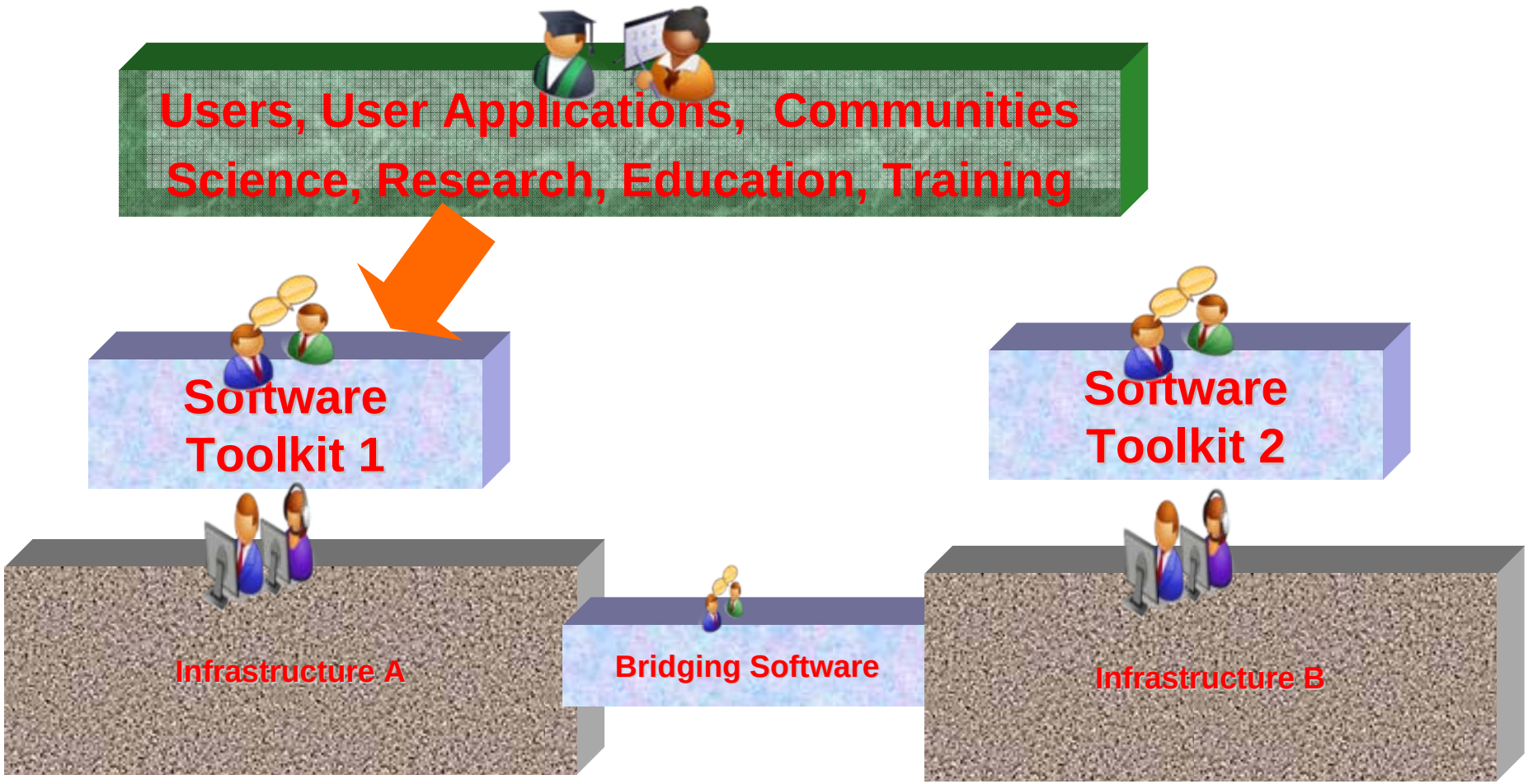
Community Common Software and Support
Science, Research, Education, Training

Software, Services & People to help
(~60 modules in the Common Virtual Data Toolkit)

Processing, Storage, Data and Administrators,
(For OSG at ~75 US Labs and Universities)

Bridging software ensures Uniformity presented to the User across Multiple Software Implementations

Open Science Grid



Software Components

Open Science Grid

Command Line User Interface

Application Services

Job Scheduling & Management

Data Handling & Management

Security

Inform-ation

Monito-ring, Error Reporting

Operating System

Why Depend on OSG for Software?

User communities have long term needs for distributed systems support and operations which can be supplied by dedicated organization.

Common software is **more efficiently** tested, packaged evolved, supported across multiple s/w developers and OS versions.

Change and evolution is managed by experts and knowledge is transferred more efficiently.

Central Security Organization helps part-time administrators, new recruits, and timely, consistnt broad response to incidents.

"Providing the security framework that promotes autonomous and open science collaboration…"

Ensuring security is not compromised because of Open Science

Ensuring science is not burdened because of security

Mine Altunay,

The OSG Security Officer. She is a Computer Scientist at Fermilab

OU is an Early Member

Many contributions

- Exemplar Cross-Campus organization and sharing of computing.
- Testing of new versions of Software
- Allowing computing and caches to be used by many different communities
- Bringing new applications to the table locally and across DOSAR sites.
- Contributions to software including accounting, metrics
- Support for Sao Paolo regional grid.

Looking forward to more to come!

Spotlight on an OSG Contributor

Horst Severini is the grid computing coordinator of the Oklahoma Center for High Energy Physics at the University of Oklahoma in Norman . He is in charge of the operation of the OU part of the US ATLAS Southwest Tier 2 Center, which also includes UT Arlington and Langston University . Besides ATLAS, the OU High Energy Physics (OUHEP) group is also a member of the D0 and DOSAR VOs within OSG.

OUHEP has several OSG CEs installed, most of which run jobs for the ATLAS, D0 and DOSAR VOs. Horst and his OUHEP colleagues also maintain an OSG testbed site, always an early adopter of new OSG integration releases, that allows them to help debug and fix integration and deployment issues. Furthermore, they work with the OSG accounting and monitoring groups on software tests.

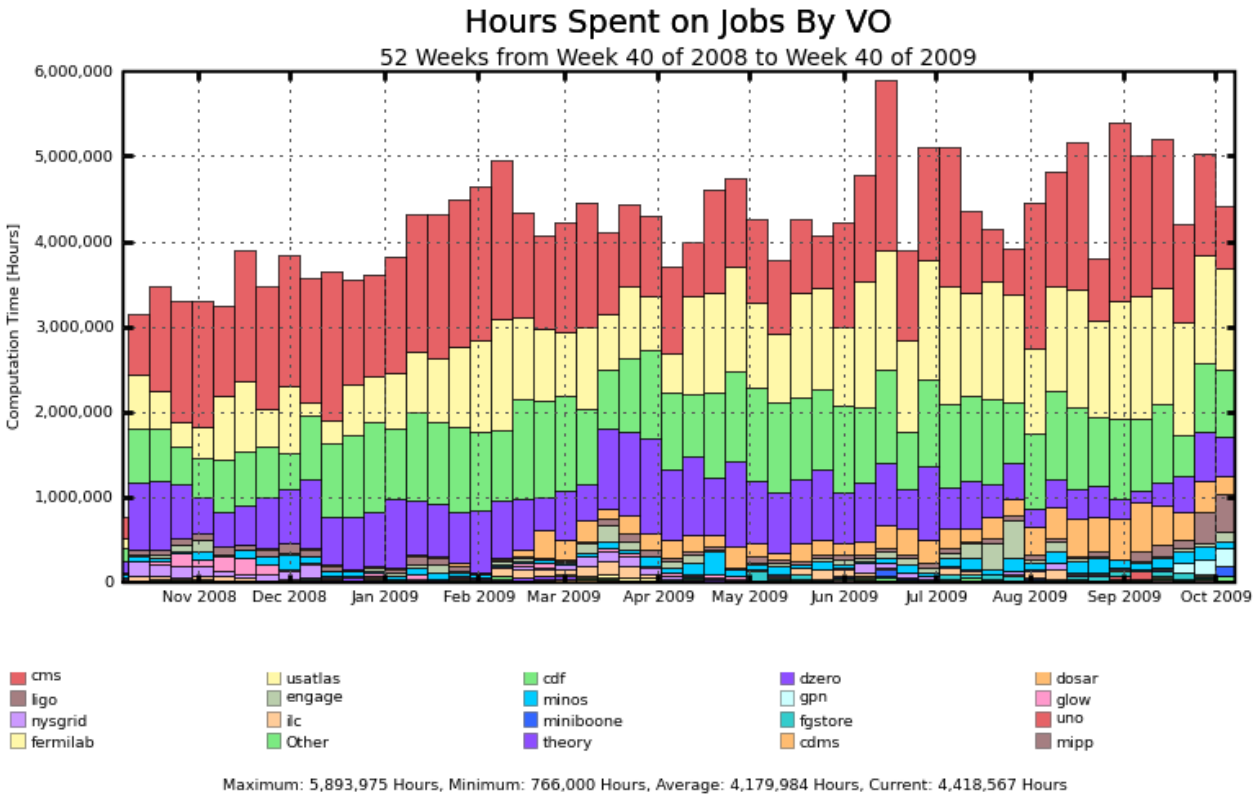
Horst is also the associate director of the OU Supercomputing Center for Education and Research, where he manages the OU Condor project. The project aims to assemble all OU campus student lab PCs into a campus

OSGNews

Nov 2006.

Support for Production Running

Across Independent Facilities for Independent User Communities relying on Production Software: Talk by Dan Fraser later today.



Knowledge Transfer: Other Campus-Regional Grids

- **Sao Paolo State Grid** in Brazil – GridUNESP – now has **7 individual sites/cluster.**
- Organizationally intending to become **Regional Grid**
- **OU contributions to transfer knowledge,** OSG services and software.
- Another **Autonomous Federated Infrastructure** that will interface to and depend on some OSG, or OSG partner, help.

“viral” extension

Grid Schools and Training sessions give hands on training in technologies and use.

“Engagement” helps early adopters get successful runs.

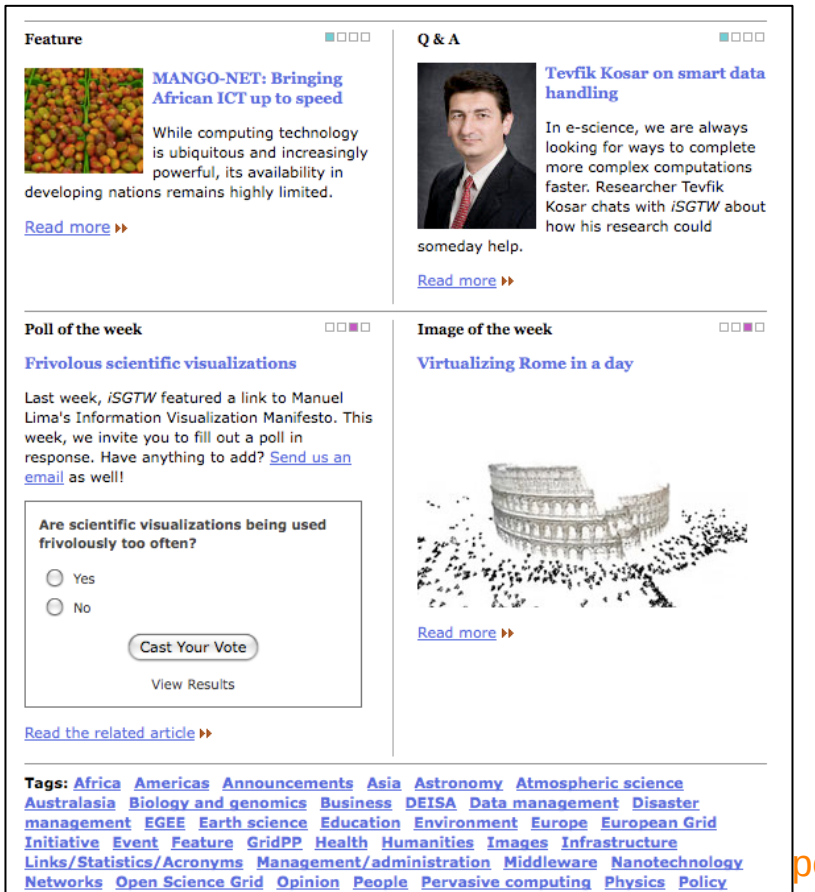
Participate in SC and TeraGrid education workshops.

Extend software through working as part of research groups themselves.

Goal to integrate into OSG program of work and increase collaboration with others.

International Science Grid This Week

Initiated by OSG, adopted by Europeans, now potentially to include TeraGrid.



Feature

MANGO-NET: Bringing African ICT up to speed

While computing technology is ubiquitous and increasingly powerful, its availability in developing nations remains highly limited.

Read more ▸▸

Q & A

Tevfik Kosar on smart data handling

In e-science, we are always looking for ways to complete more complex computations faster. Researcher Tevfik Kosar chats with *iSGTW* about how his research could someday help.

Read more ▸▸

Poll of the week

Frivolous scientific visualizations

Last week, *iSGTW* featured a link to Manuel Lima's Information Visualization Manifesto. This week, we invite you to fill out a poll in response. Have anything to add? Send us an email as well!

Are scientific visualizations being used frivolously too often?

- ○ Yes
- ○ No

Cast Your Vote

View Results

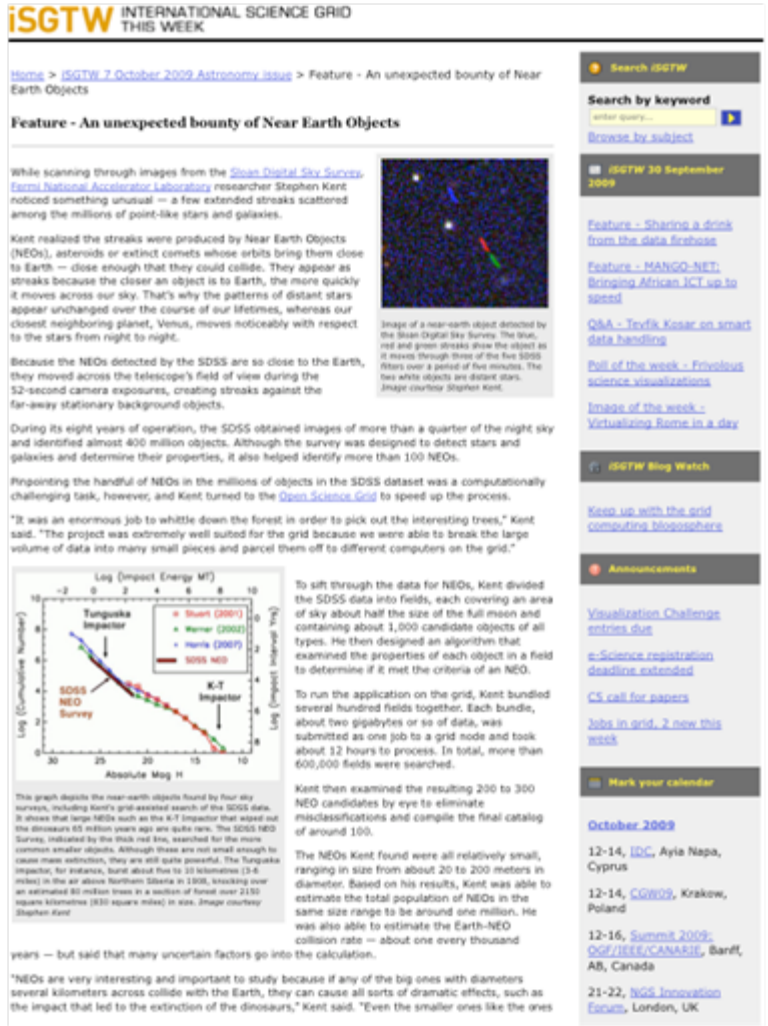
Read the related article ▸▸

Image of the week

Virtualizing Rome in a day

Read more ▸▸

Tags: Africa Americas Announcements Asia Astronomy Atmospheric science Australasia Biology and genomics Business DEISA Data management Disaster management EGEE Earth science Education Environment Europe European Grid Initiative Event Feature GridPP Health Humanities Images Infrastructure Links/Statistics/Acronyms Management/administration Middleware Nanotechnology Networks Open Science Grid Opinion People Pervasive computing Physics Policy



iSGTW INTERNATIONAL SCIENCE GRID THIS WEEK

Home > iSGTW 7 October 2009 Astronomy issue > Feature - An unexpected bounty of Near Earth Objects

Feature - An unexpected bounty of Near Earth Objects

While scanning through images from the Sloan Digital Sky Survey, Fermi National Accelerator Laboratory researcher Stephen Kent noticed something unusual — a few extended streaks scattered among the millions of point-like stars and galaxies.

Kent realized the streaks were produced by Near Earth Objects (NEOs), asteroids or extinct comets whose orbits bring them close to Earth — close enough that they could collide. They appear as streaks because the closer an object is to Earth, the more quickly it moves across our sky. That's why the patterns of distant stars appear unchanged over the course of our lifetimes, whereas our closest neighboring planet, Venus, moves noticeably with respect to the stars from night to night.

Because the NEOs detected by the SDSS are so close to the Earth, they moved across the telescope's field of view during the 52-second camera exposures, creating streaks against the far-away stationary background objects.

During its eight years of operation, the SDSS obtained images of more than a quarter of the night sky and identified almost 400 million objects. Although the survey was designed to detect stars and galaxies and determine their properties, it also helped identify more than 100 NEOs.

Pinpointing the handful of NEOs in the millions of objects in the SDSS dataset was a computationally challenging task, however, and Kent turned to the Open Science Grid to speed up the process.

"It was an enormous job to whittle down the forest in order to pick out the interesting trees," Kent said. "The project was extremely well suited for the grid because we were able to break the large volume of data into many small pieces and parcel them off to different computers on the grid."

To sift through the data for NEOs, Kent divided the SDSS data into fields, each covering an area of sky about half the size of the full moon and containing about 1,000 candidate objects of all types. He then designed an algorithm that examined the properties of each object in a field to determine if it met the criteria of an NEO.

To run the application on the grid, Kent bundled several hundred fields together. Each bundle, about two gigabytes or so of data, was submitted as one job to a grid node and took about 12 hours to process. In total, more than 600,000 fields were searched.

Kent then examined the resulting 200 to 300 NEO candidates by eye to eliminate misclassifications and compile the final catalog of around 100.

The NEOs Kent found were all relatively small, ranging in size from about 20 to 200 meters in diameter. Based on his results, Kent was able to estimate the total population of NEOs in the same size range to be around one million. He was also able to estimate the Earth-NEO collision rate — about one every thousand years — but said that many uncertain factors go into the calculation.

"NEOs are very interesting and important to study because if any of the big ones with diameters several kilometers across collide with the Earth, they can cause all sorts of dramatic effects, such as the impact that led to the extinction of the dinosaurs," Kent said. "Even the smaller ones like the ones

Search iSGTW

Search by keyword

Browse by subject

iSGTW 30 September 2009

Feature - Sharing a drink from the data firehose

Feature - MANGO-NET: Bringing African ICT up to speed

Q&A - Tevfik Kosar on smart data handling

Poll of the week - Frivolous science visualizations

Image of the week - Virtualizing Rome in a day

iSGTW Blog Watch

Keep up with the grid computing blogosphere

Announcements

Visualization Challenge entries due

e-Science registration deadline extended

CS call for papers

Jobs in grid, 2 new this week

Mark your calendar

October 2009

12-14, IDC, Ayia Napa, Cyprus

12-14, CGW09, Krakow, Poland

12-16, Summit 2009: OGF/IEEE/CANARIE, Banff, AB, Canada

21-22, NGS Innovation Forum, London, UK

Futures – Vision

Seeding Expansion in University Campus Grids

~100 US LHC Tier-3s will participate to get data through OSG and run jobs locally and remotely.

New Neescomm (Earthquake shake tables) Community Collaboration sub-committee. 14 equipment sites collect, archive and share data.

Internet2, Educause educate the CIOs, CFOs and Presidents through workshops and “CIDays”.

OSG to participate with TeraGrid Campus Champions.

…… Your ideas ??

The OSG & TeraGrid Mix

- OSG and TeraGrid use same fundamental software.
  - Client/user-side software should be easily adaptable to use both.
  - Several portals access both (Nanohub, RENCI, GIGI)
- To use TeraGrid you need to apply for and be granted an Allocation.
- To use OSG you need to register with a Community or VO. OSG provides some “ready made” VOs:
  - OSGEDU is a VO for Students and Teachers
  - Engage is a VO where we help you get your code to run and then run “helper services” for you to monitor your runs, select sites where your code will run etc.

A National Vision

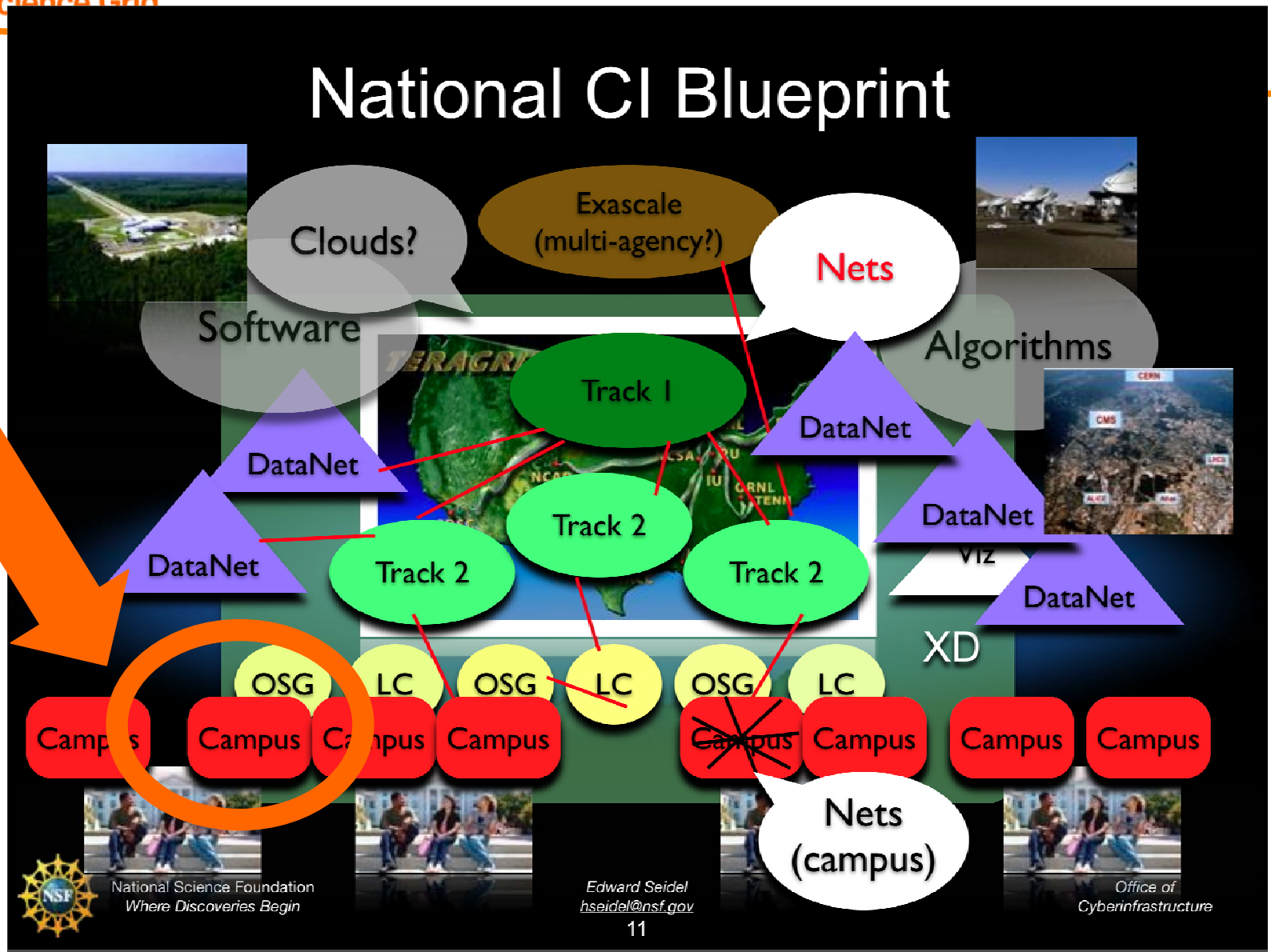
We (OSG) aim to be a driver for and

An integral part of the

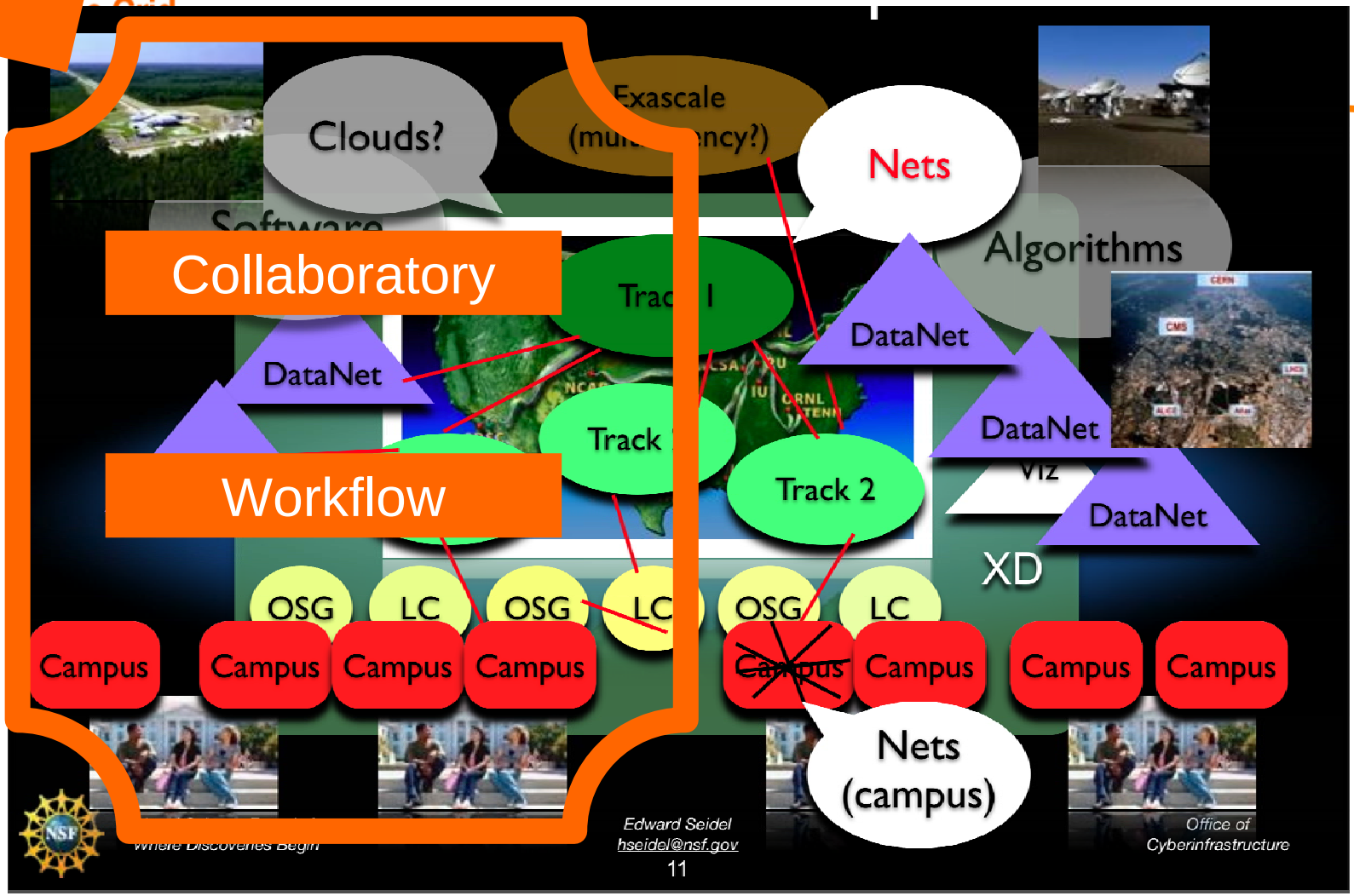
National Cyber-Infrastructure

! With Continued Help and Involvement !

National Science Foundation Landscape – Ed Seidel, Feb 2009



OSG looks at the Whole System



Monday, February 2, 2009

E.g. Smooth support for all scales of parallel jobs

New OSG satellite project to test small-scale parallel jobs that use multiple cores on one CPU-box.

Fills gap between single processor, high throughput, support from OSG and large MPI, high performance, from TeraGrid.

OSCER to be the first site used!

More OSG-TG Collaboration coming? Note words in TG extension project:

* Continue to support several resources into 2011. This includes the Track 2 systems, ….
* Allow high-throughput, Open Science Grid-style jobs.
* Enable exploration of interoperability and technology sharing.
* Provide a transition platform for users coming from university- or departmental-level resources.
* Support unique compute platforms and massive storage systems….

Provide software Glue for use of the full suite of computing resources out there

Integrate commercial Cloud computing –Amazon, Google etc.

"No campus left behind".

Support Virtual Machines of various types.

Bringing in the next generaions.

Unique HPC & visualization engines.

Integrate the laptops,iphones..

Shared data repositories & large local data silos.

Data from multi-robots and distributed sensors.