Open Science Grid

XSEDE

Extreme Science and Engineering
Discovery Environment

OSG and TeraGrid
A Comparison

Alain Roy

OSG Software Coordinator

U. Wisconsin, Madison

John-Paul “J-P” Navarro

TeraGrid GIG Area Co-Director for Software Integration

U. of Chicago / Argonne National Laboratory

OSG Summer School

June 30, 2011

Open Science Grid

XSEDE

Extreme Science and Engineering Discovery Environment

OSG and ~~TeraGrid~~/XSEDE A Comparison

Alain Roy
OSG Software Coordinator
U. Wisconsin, Madison

John-Paul “J-P” Navarro
XSEDE Software Development & Integration Deputy Manager
U. of Chicago, Argonne National Laboratory

OSG Summer School
June 30, 2011

Open Science Grid

Introduction

XSEDE
Extreme Science and Engineering Discovery Environment

OSG Mission statement

The Open Science Grid aims to promote discovery and collaboration in data-intensive research by providing a computing facility and services that integrate distributed, reliable and shared resources to support computation at all scales.

OSG funding agencies: NSF and DOE

XSEDE Vision statement

Enhance the productivity of scientists and engineers by providing them with new and innovative capabilities, and thus facilitate scientific discovery while enabling transformational science/engineering and innovative educational programs.

TeraGrid funding agencies: NSF

What is OSG?

- You' ve seen a lot about OSG in the last few days, so hopefully you have a good idea of what OSG is.

  A few reminders:

  – 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.

  – OSG integrates computing and storage resources from over 80 sites in the U.S. and beyond

What is XSEDE?

- Is a distributed “grid” facility, managed like a large cooperative research group, not a consortium:
  - One coordinating XSEDE award (little hardware, backbone network)
  - ~10 initial evolving service provider “SP” infrastructure awards
- XSEDE coordination award provides single:
  - Portal, sign-on, help desk, allocations process, advanced user support, EOT, campus champions, coordinated software, integration process
- The SP awards include:
  - Large tightly coupled distributed memory clusters
    - Ranger (TACC) 579 TF, Kraken (NICS) > 1 PF
  - Large share memory machines
  - Clusters with Infiniband; Condor pools
  - Viz, Storage, and experimental (FutureGrid) resources
- Targets US based open science research (NSF primarily)

XSEDE’s Distinguishing Characteristics

- Foundation for a national cyberinfrastructure ecosystem
  - comprehensive suite of advanced digital services will federate with other high-end facilities and campus-based resources
- Unprecedented integration of **diverse** digital resources
  - innovative, open architecture making possible the continuous addition of new technology capabilities and services

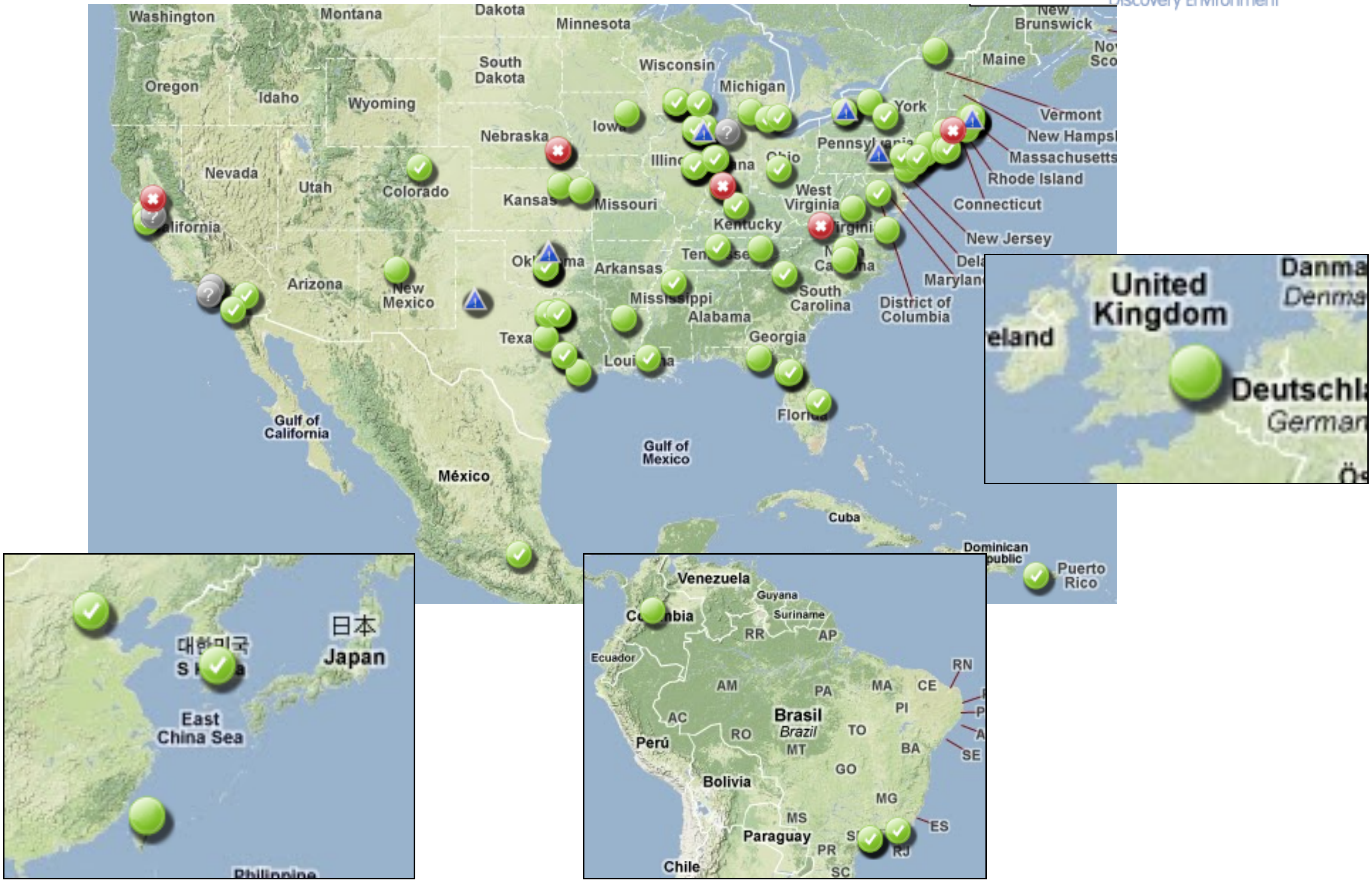
XSEDE's Distinguishing Characteristics – User Services

- Unified and comprehensive user support services
  - designed to advance science and engineering
  - provided by experts in the application of technology
  - new and expanded facets of advanced support
    - external shorter-term contracting for expertise beyond that in current team; Novel and Innovative Projects
- More mature operational practices
  - increased focus on productivity and ease of use
  - increased and enhanced security, reliability, and quality assurance

Open Science Grid

Some of the OSG Sites

XSEDE
Extreme Science and Engineering
Discovery Environment

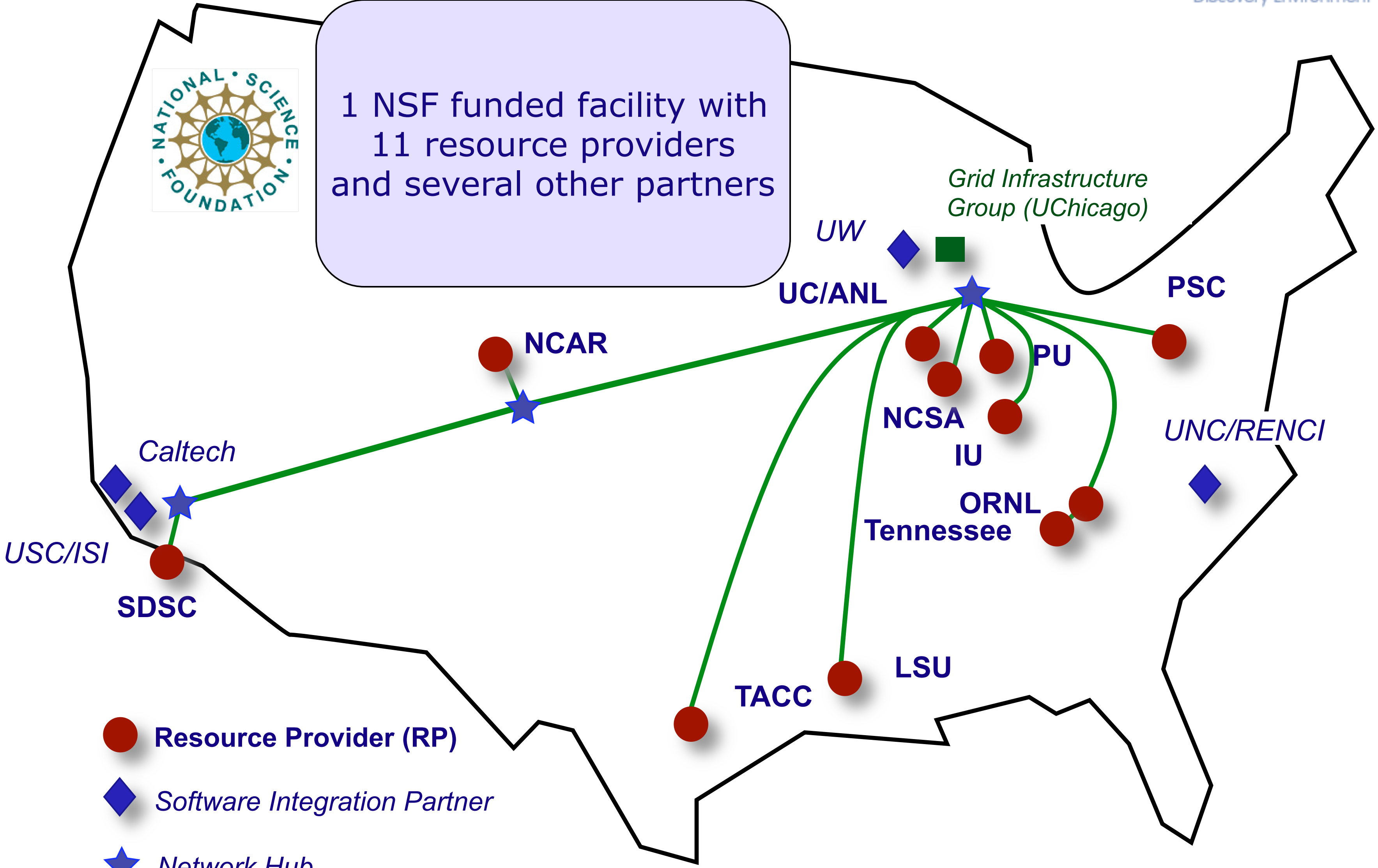


Open Science Grid

The TeraGrid Partners

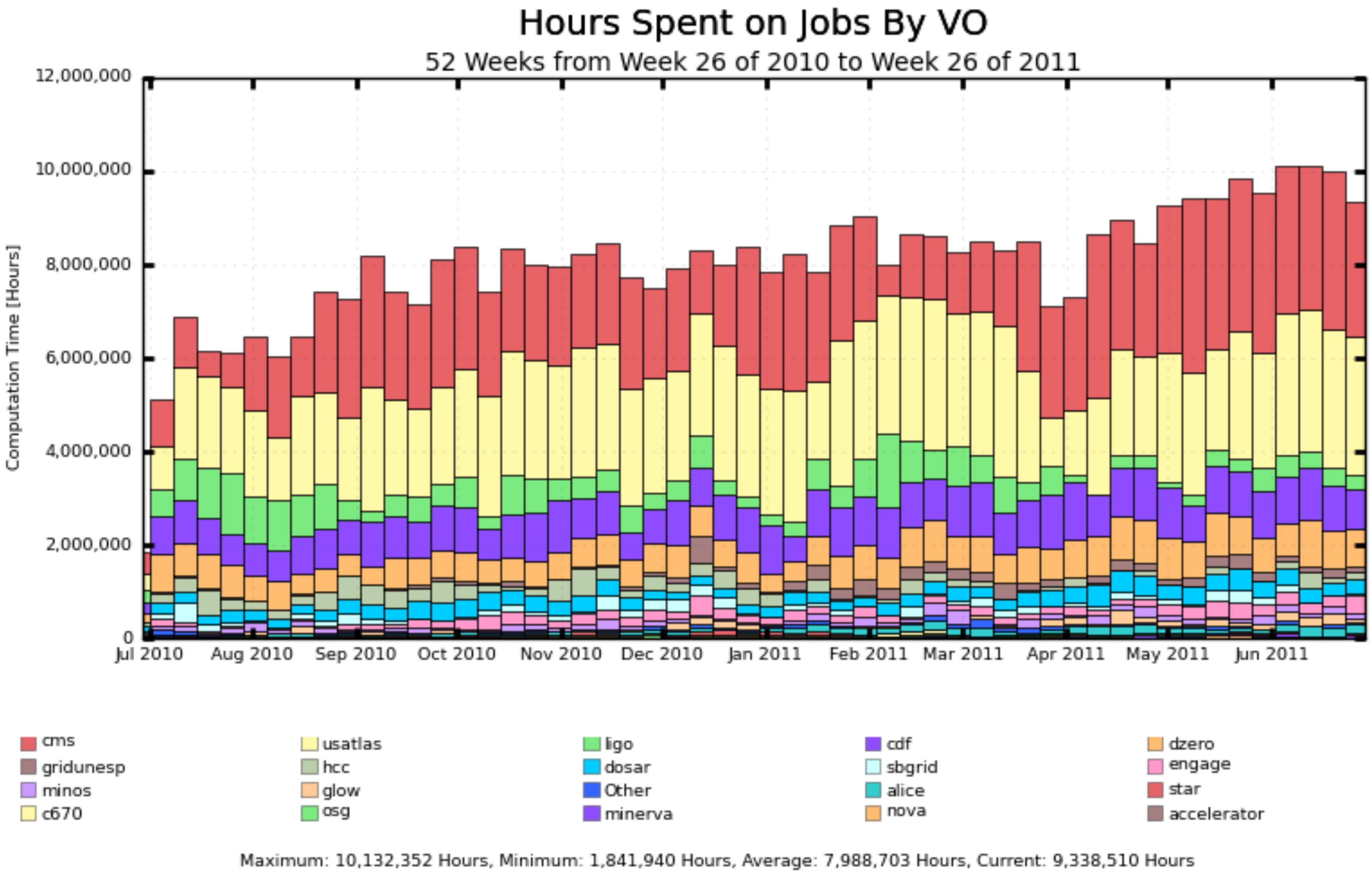
XSEDE
Extreme Science and Engineering
Discovery Environment

1 NSF funded facility with 11 resource providers and several other partners



- ● Resource Provider (RP)
- ◆ *Software Integration Partner*
- ★ *Network Hub*

Some OSG Job stats



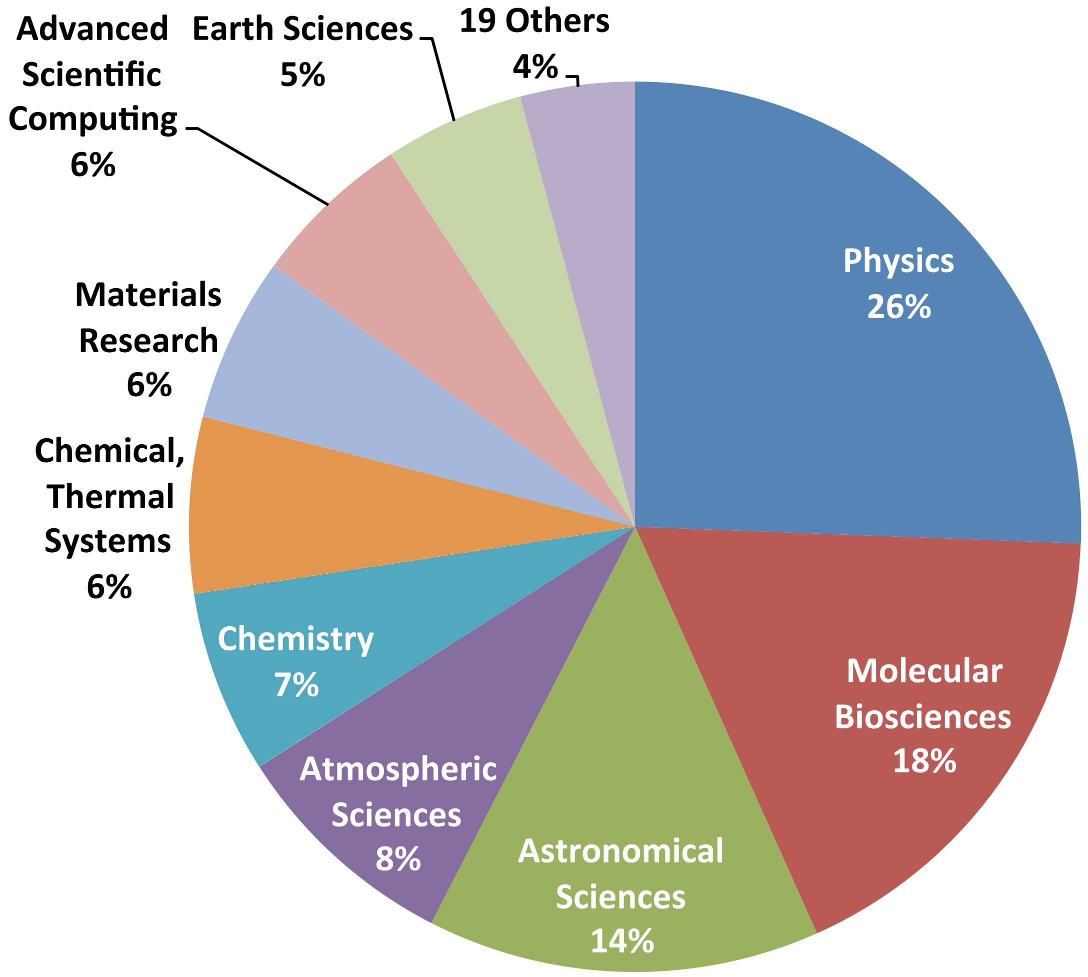
Average: ~3,500,000 jobs/week

TeraGrid Usage

Open Science Grid

XSEDE

Extreme Science and Engineering Discovery Environment

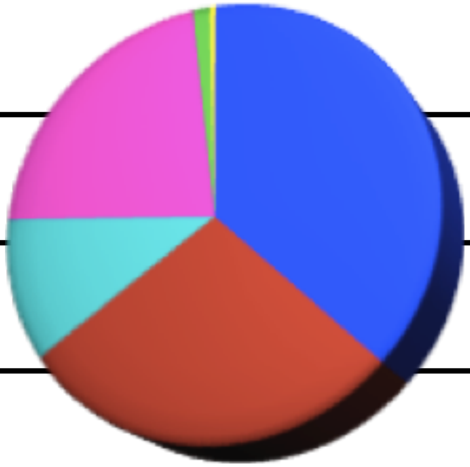


>27B NUs Delivered in 2009
Over 1,200 projects supported

Open Science Grid

How TeraGrid Is Used

XSEDE
Extreme Science and Engineering
Discovery Environment

Use Modality	Community Size (rough est. - number of users)
Batch Computing on Individual Resources	850
Exploratory and Application Porting	 650
Workflow, Ensemble, and Parameter Sweep	250
Science Gateway Access	500
Remote Interactive Steering and Visualization	35
Tightly-Coupled Distributed Computation	10

2006 data

Open Science Grid

TeraGrid & OSG Resources

XSEDE

Extreme Science and Engineering Discovery Environment

	OSG	TeraGrid
HPC Resources	Some have Myrinet/Infiniband	12
HPC TeraFlops		>2 PF
HPC Processor		>200K CPUs
HTC Resources	~93 compute elements	4
HTC Processors	421M CPU hours/year Typical ~70K CPUs in use	>27K CPUs
Storage	~31 PB	3 PB disk + 60 PB tape
Network	varies, shared Internet2	dedicated 10 Gbps
Visualization		4 resources
Other		VMs, GPUs, FPGAs

OSG Software

OSG Software Stack (a.k.a VDT)

- Set of software components shared across OSG
- Built, packaged, tested, and distributed as coherent software distribution

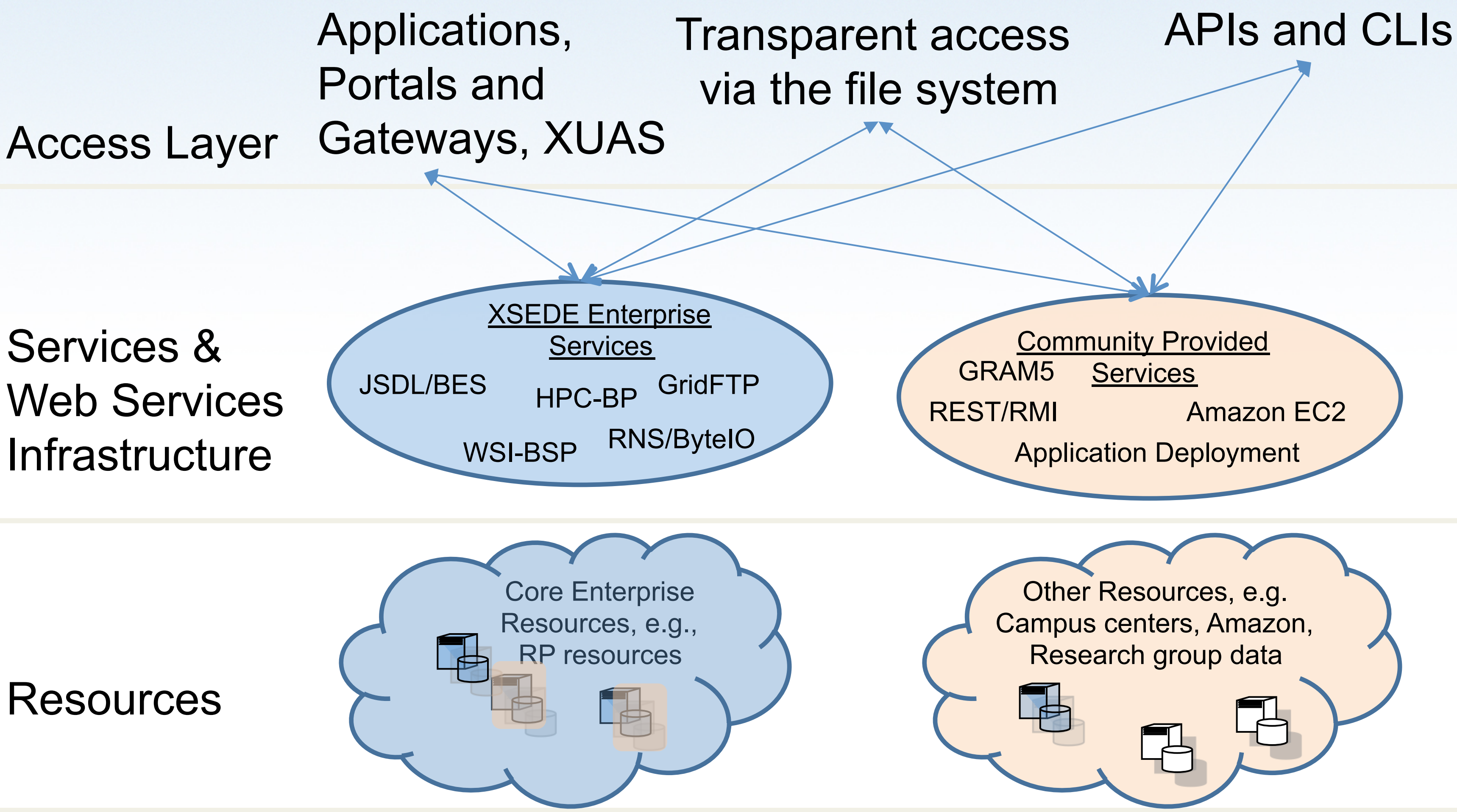
Major components

- Compute Element
  - Accepts jobs to site
  - Has accounting, authentication/authorization, information services, and utilities
- Storage Element
  - Manages storage at site
  - Has accounting, authentication/authorization, information services, and utilities
- Worker Node
  - Software required at each worker node (basic client tools…)
- Client
  - Software used to submit to OSG
  - Condor-G, data clients, security clients, etc…

Distinguishing characteristics: Architecture

- XSEDE is *designed* for innovation and evolution
  - there *is* an architecture defined
    - based on set of design principles
    - rooted in the judicious use of standards and best practices
- Professional systems engineering approach
  - responds to evolving needs of existing, emerging, and new communities
    - incremental development/deployment model
  - new requirements gathering processes
    - ticket mining, focus groups, usability panels, shoulder surfing
  - ensure robustness and security while incorporating new and improved technologies and services
  - process control, quality assurance, baseline management, stakeholder involvement

XSEDE Architecture



Access Layer

Applications, Portals and Gateways, XUAS

Transparent access via the file system

APIs and CLIs

Services & Web Services Infrastructure

XSEDE Enterprise Services

JSDL/BES

HPC-BP

GridFTP

WSI-BSP

RNS/ByteIO

Community Provided Services

GRAM5

REST/RMI

Amazon EC2

Application Deployment

Resources

Core Enterprise Resources, e.g., RP resources

Other Resources, e.g. Campus centers, Amazon, Research group data

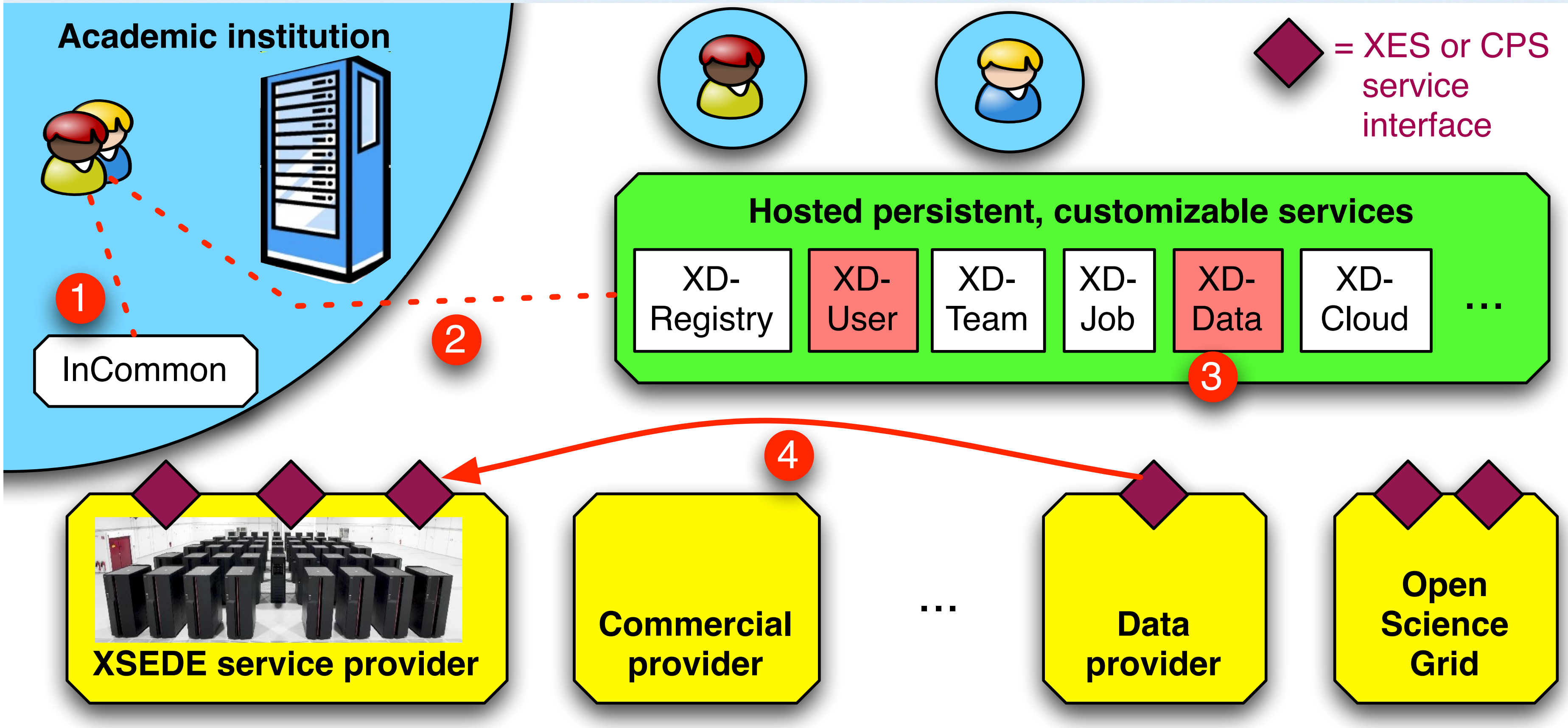
Access Layer – XUAS, Portals, Gateways, and GUIs

- XUAS – XSEDE User Access Services
  - Web browser/GUI, RESTfull API, and CLI
  - Globus Online hosted services
    - XD-Data, (XD-User, XD-Team, XD-Job, etc..)
- XSEDE client tools
  - Graphical User Interface
    - JSDL tool, queue manager, shell, access control tool, browser, job management, etc.
- Portal(s), plus Gateways accessing XES and CPS as necessary

Access Layer – APIs and CLIs

- SAGA – Standard API for Grid Applications
- Back-end specific APIs (**not recommended**)
  - Genesis II
  - Globus
  - UNICORE 6
- GSI-SSH
- Grid shell (CLI) that interacts with standard back-ends
  - qsub, qstat, qkill, qcomplete, run, create-queue, export a directory

XSEDE User Access Services “XUAS”



Academic institution

InCommon

= XES or CPS service interface

Hosted persistent, customizable services

XD-Registry

XD-User

XD-Team

XD-Job

XD-Data

XD-Cloud

…

XSEDE service provider

Commercial provider

…

Data provider

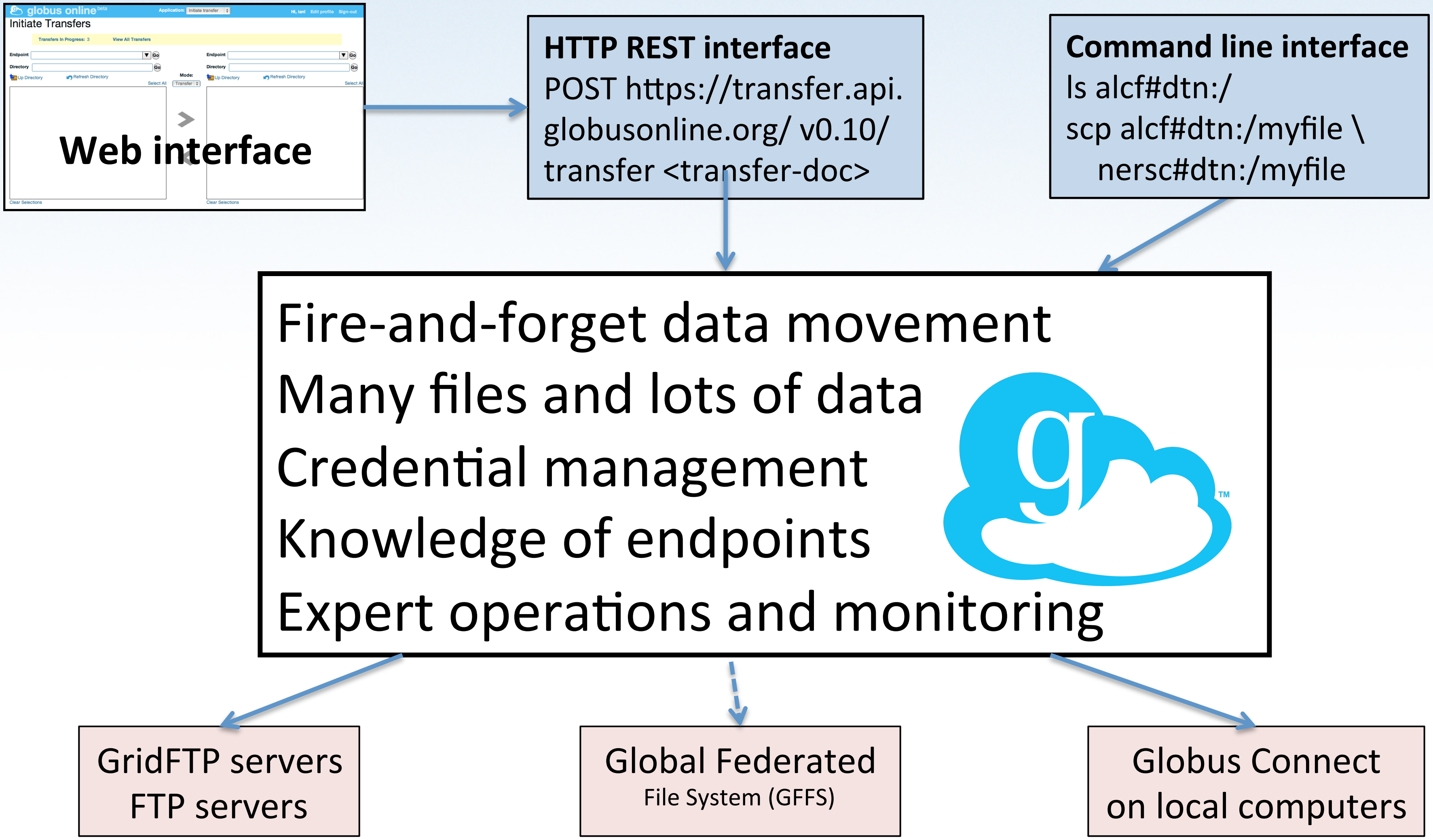
Open Science Grid

— User-facing perspective on cyberinfrastructure

— Integrated view of XSEDE and non-XSEDE (e.g., campus) resources

— One-stop responsibility for troubleshooting and support

E.g.: Globus Online data mover (XD-Data)



Web interface

HTTP REST interface
POST https://transfer.api.globusonline.org/ v0.10/ transfer <transfer-doc>

Command line interface
ls alcf#dtn:/
scp alcf#dtn:/myfile \
nersc#dtn:/myfile

Fire-and-forget data movement
Many files and lots of data
Credential management
Knowledge of endpoints
Expert operations and monitoring

GridFTP servers
FTP servers

Global Federated
File System (GFFS)

Globus Connect
on local computers

Access Layer – XSEDE Wide File System XWFS

- Based on GPFS or Lustre-WAN
  - “Trade study” during year one
    - Involves requirements gathering, testing, consideration of quality attributes such as reliability, performance, cost, down-stream risk
- Mounted at all NSF-funded sites
- Requires a high degree of trust

Access Layer – Global Federated File System GFFS

- Spans centers, campuses, individual labs
- Maps resources (files, databases, archives, clusters, supercomputers, running jobs) into a global namespace
  - Research groups can map a directory tree on a lab machine directly into the shared name space
  - Research groups can map their PBS controlled cluster directly into the shared name space
- Maps global namespace into local file system, e.g., /home/grimshaw/grid
- Resources can be accessed using file system commands and POSIX IO, e.g., ls, cat, vi, cp, thus you can
  - Access (read) data as it comes off an instrument in a colleagues lab in another university
  - ***cd*** into the working directory of a running job and view and edit files without knowing where the job is running
  - ***cp*** a job description file into a directory representing a compute resource to execute the job
  - ***ls*** a compute resource to see running, queued, and finished jobs

Security Comparison

OSG security implementation:

- X.509 certificates, assigned to and managed by users
- VOMS to:
  - Manage VO membership
  - Extend X.509 certificate with VO membership/role

TeraGrid security implementation:

- X.509 certificates, assigned to and managed by users
- Custom project/VO membership system (tgcdb)
- MyProxy credential management and repository
- Science gateway community accounts w/ GridShib user attributes

XSEDE security implementation:

- InCommon authentication, SAML
- XES security model based on WS-Security, WSI-BSP, WS-Trust
- XD-User, XD-Team project/VO membership management

Information Discovery In OSG

- Information collected at every Compute/Storage Element (at every site) at regular intervals
  - Information conforms to schema agreed upon with international collaborators.
- Information forwarded to two services
  - BDII: Centralized LDAP server (particularly used for LHC experiments)
  - RESS: Condor collector (basis for OSG matchmaker)

Information Discovery In TeraGrid

- Service Providers publish:
  - Capability, compute info, resource info, local s/w
  - GLUE2 schema and other custom schemas
- TeraGrid aggregates in central Integrated Information Services "IIS"
  - Plus some additional central database information
  - Available via REST (XML, JSON), and ~~Globus MDS4 (XML)~~
  - More information: http://info.teragrid.org/

Information Discovery In XSEDE

- Evolves TeraGrid's Information Services
- Introduces new capabilities
  - AMQP publish/subscribe
  - Can describe all XSEDE integrated digital products: resources, services, software, data collections, training products, events, cloud appliances, etc..
  - **Digital Product Advisor**

Job Submission Comparison

OSG Job Submission

- Common infrastructure:
  - Globus GRAM accepts jobs at a site
  - Condor-G submits jobs to a site
- Workload management systems on top of this
  - OSG Matchmaker (Condor-G with matchmaking based on OSG information service)
  - Glideins/Pilot jobs

XSEDE Job Submission

- XES: Unicore 6, GFFS
- CPS: GRAM5, Genesis II BES
- Local login node, Condor-G
- Condor-Matchmaker, Genesis Metascheduler
- Science Gateway based submission
- Co-scheduler, Advanced reservation support

Open Science Grid

Data Access Comparison

XSEDE

Extreme Science and Engineering Discovery Environment

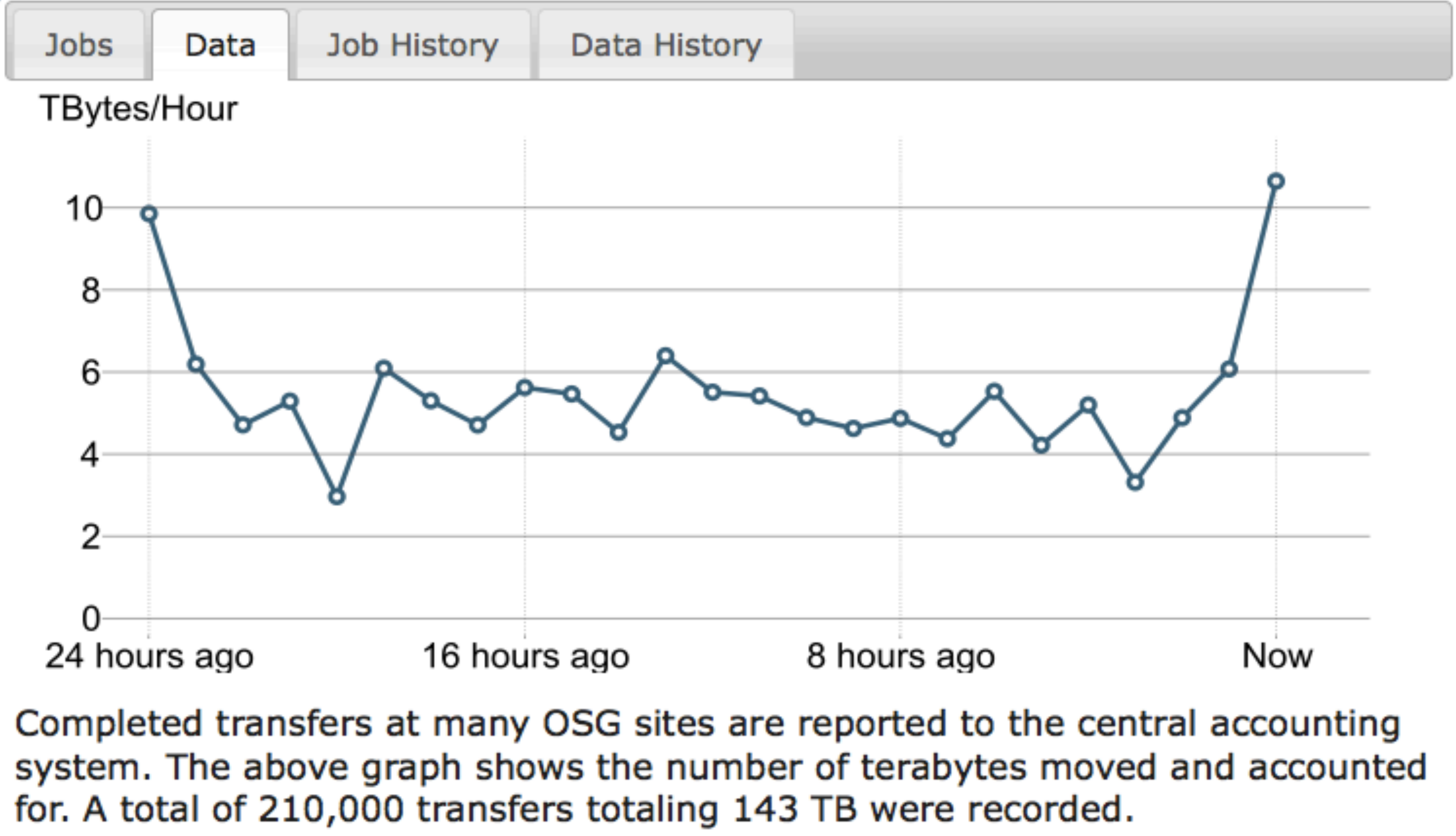
- OSG data management systems:
  - GridFTP for data transfer
  - OSG_APP_DATA/OSG_DATA for minimalist storage.
  - Storage Elements with SRM data better storage management.
  - Some systems built on top of this: FTS, PhedEx, DQ2.
- XSEDE data management systems:
  - GridFTP for data transfer, also GSI OpenSSH w/ HPN
  - XUAS GO-Data movement
  - Wide-area Lustre and GPFS file-systems (evaluation)
  - Data replication, data management (evaluation)
  - GFFS – Globus Federated File-System

Open Science Grid

Some OSG data stats

XSEDE

Extreme Science and Engineering Discovery Environment



Completed transfers at many OSG sites are reported to the central accounting system. The above graph shows the number of terabytes moved and accounted for. A total of 210,000 transfers totaling 143 TB were recorded.

- More pretty graphs at:
  - http://display.grid.iu.edu/