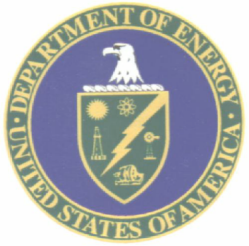


Open Science Grid

at Condor Week

Ruth Pordes
Fermilab
April 25th 2006

Outline

- OSG goals and organization
- Drivers and use today
- Middleware
- Focus and roadmap

What, Who is Open Science Grid?

- **High Throughput Distributed Facility**
  - Shared opportunistic access to existing clusters, storage and networks.
  - Owner controlled resources and usage policies.
- **Supports Science**
  - Funded by NSF and DOE projects.
  - Common technologies & cyber-infrastructure.
- **Open & Inclusive**
  - Collaboration of users, developers, grid technologists & facility administrators.
  - Training & help for administrators and users
  - Partner with peer grid projects.
- **Heterogeneous**

OSG Organization (& Condor project)

Executive Director:	Ruth Pordes
Facility Coordinator:	Miron Livny
Application Coordinators:	Torre Wenaus & fkw
Resource Managers:	P. Avery & A. Lazzarini
Education Coordinator:	Mike Wilde
Engagement Coord.:	Alan Blatecky
Middleware Coord.:	Alain Roy
Ops Coordinator:	Leigh Grundhoefer
Security Officer:	Don Petravick
Liaison to EGEE:	John Huth
Liaison to Teragrid:	Mark Green
Council Chair:	Bill Kramer

& 5 year proposal submitted

OSG Drivers:

- **Research groups transitioning from & extending (legacy) systems to Grids:**
- **US LHC Collaborations**
  - Contribute to & depend on milestones, functionality, capacity of OSG.
  - Commitment to general solutions, sharing resources & technologies;
- **Application Computer Scientists**
  - Real life use of technology, integration, operation.
- **Federations with Campus Grids**
  - Bridge & interface Local & Wide Area Grids.
- **Interoperation & partnerships with national/ international infrastructures**
  - Ensure transparent and ubiquitous access.
  - Work towards standards.

LIGO- gravitational wave physics;
STAR - nuclear physics, CDF, D0, - high energy physics,
SDSS - astrophysics
GADU - bioinformatics
Nanohub

NMI, Condor, Globus, SRM

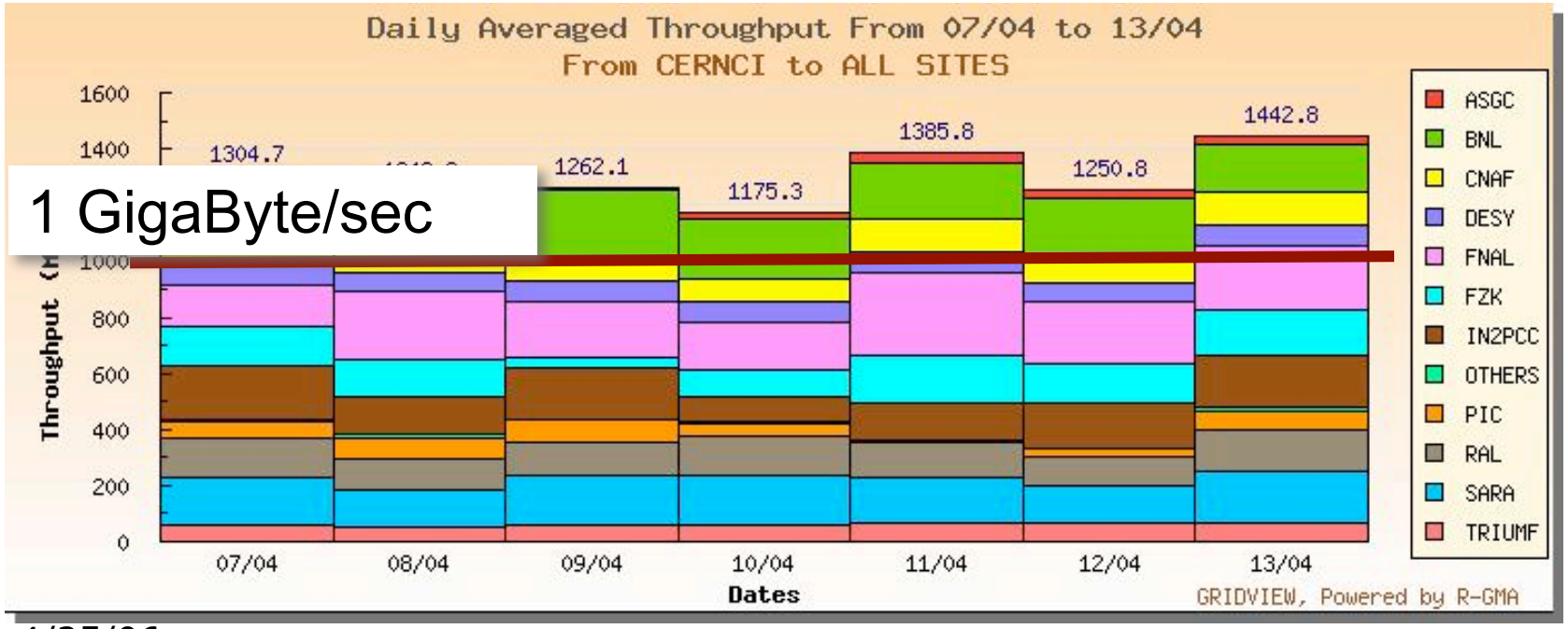
GLOW, FermiGrid, GROW, Crimson, TIGRE

EGEE, TeraGrid, INFNGrid

LHC Physics drive schedule and performance envelope

Open Science Grid

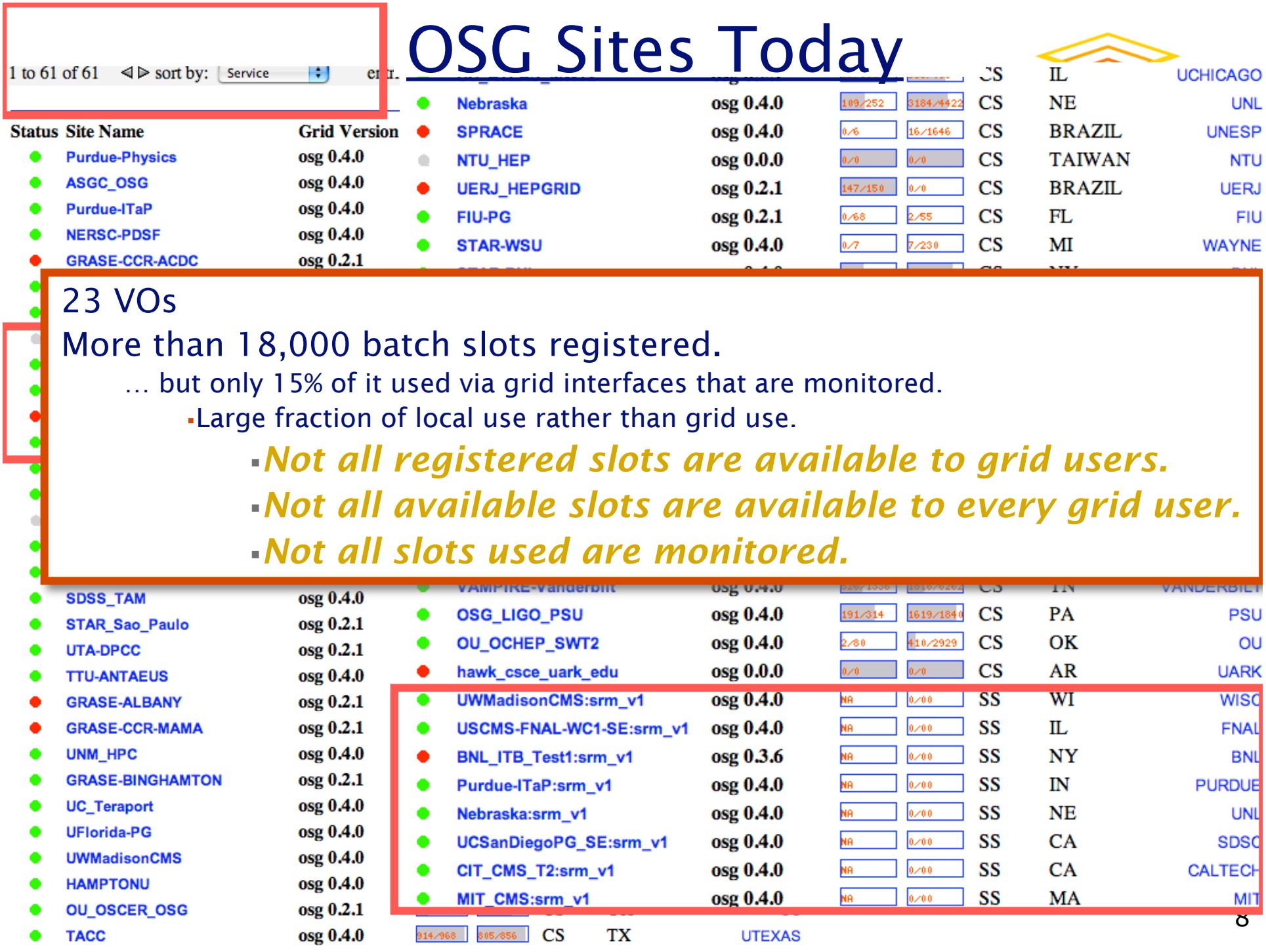
- Beam starts in 2008:
- Distributed System must serve 20PB of data in served across 30PB disk distributed across 100 sites worldwide to be analyzed by 100MSpecInt2000 of CPU.
- Service Challenges give steps to full system



Priority to many other stakeholders

- New science enabled by opportunistic use of resources
  - *E.g. From OSG Proposal:* LIGO : With an annual science run of data collected at roughly a terabyte of raw data per day, this will be critical to the goal of transparently carrying out **LIGO data analysis** on the opportunistic cycles available on other VOs hardware
- Opportunity to share use of "standing army" of resources
  - *E.g. From OSG news:* **Genome Analysis and Database Update system**, uses grid resources to analyse all publicly available genome sequence data used by over 2,400 researchers worldwide.
- Interface existing computing and storage facilities and Campus Grids to a common infrastructure.
  - *E.g. FermiGrid Strategy:* To allow opportunistic use of otherwise dedicated resources. To **save effort by implementing shared services.** To work coherently to move all of our applications and services to run on the Grid.

OSG Sites Today



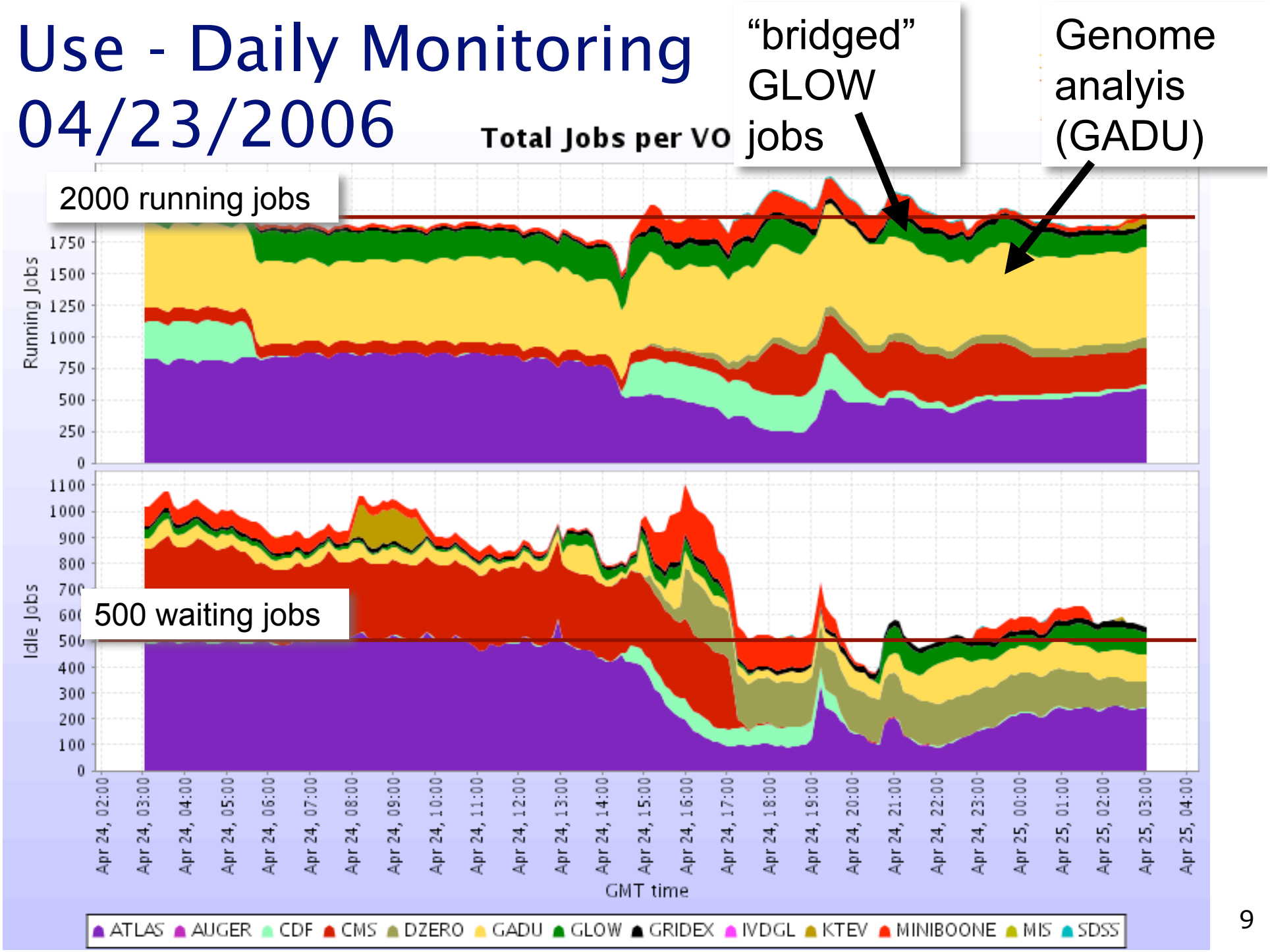
23 VOs

More than 18,000 batch slots registered.

… but only 15% of it used via grid interfaces that are monitored.

- Large fraction of local use rather than grid use.
- ***Not all registered slots are available to grid users.***
- ***Not all available slots are available to every grid user.***
- ***Not all slots used are monitored.***

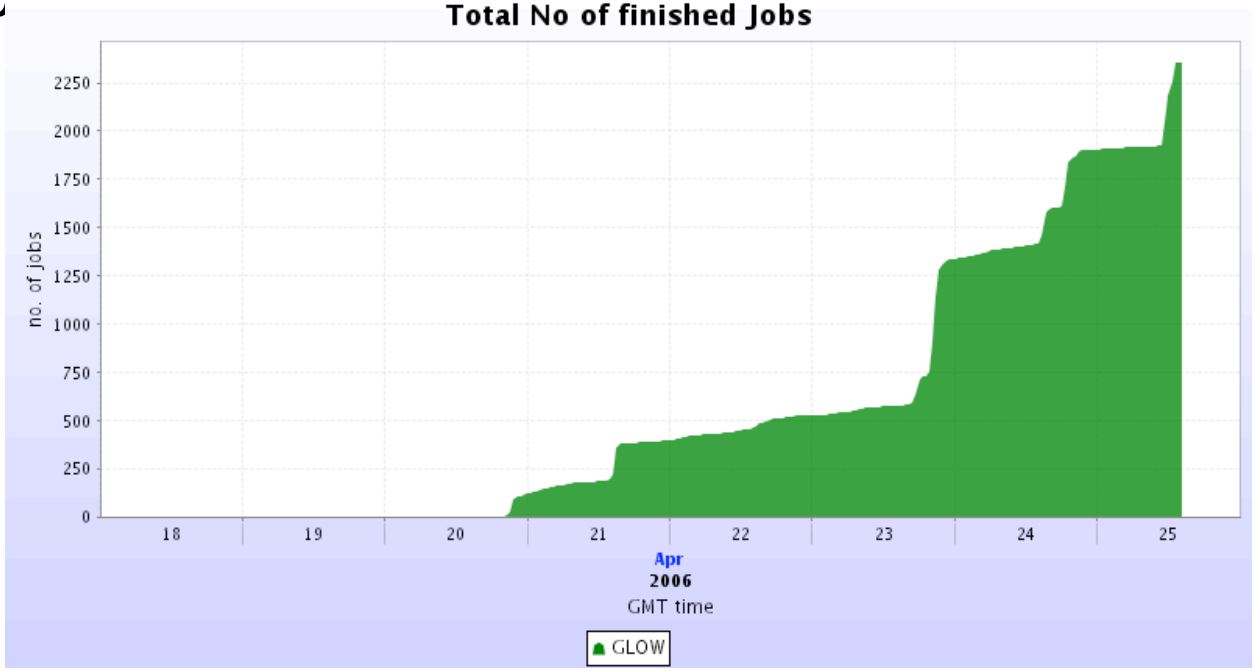
Use - Daily Monitoring 04/23/2006



Bridging Campus Grid Jobs - GLOW

Open Science Grid

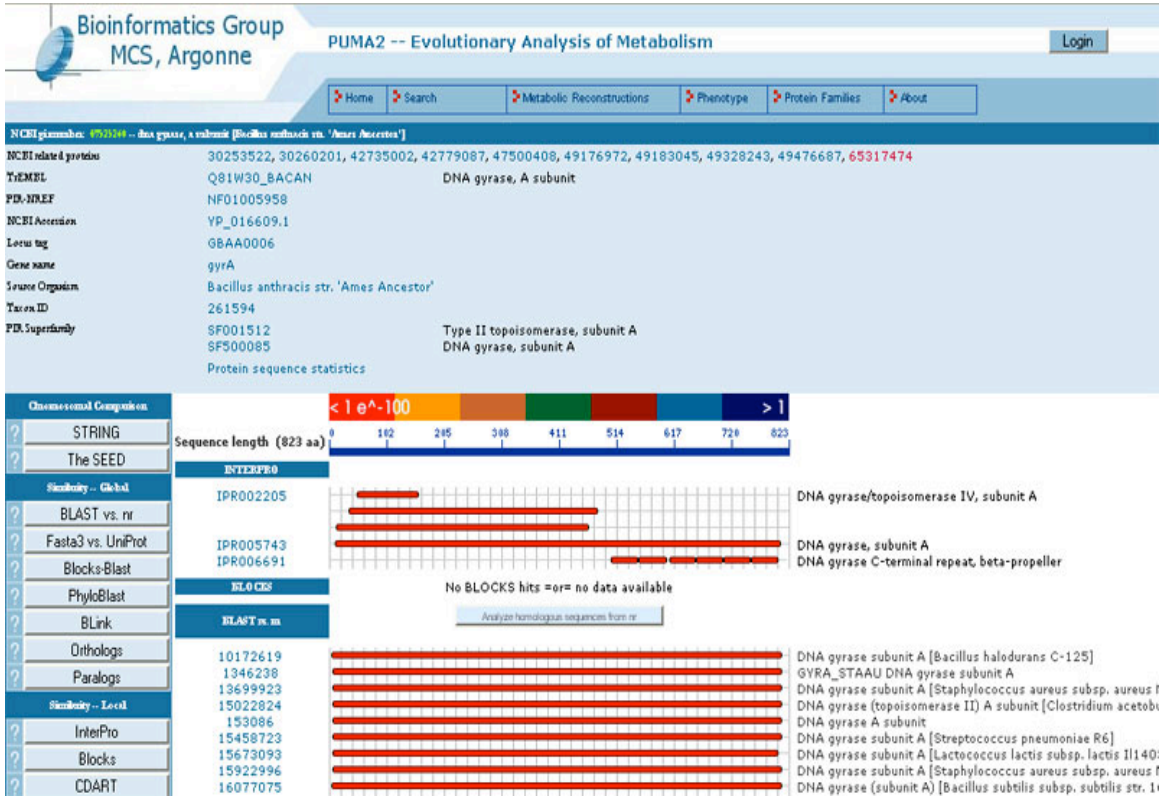
- Dispatch jobs from local security, job, storage infrastructure and “uploading” to wide-area infrastructure.
- Fast ramp up in last week.
- Currently running the **football pool problem** which has application in data compression, coding theory, and statistical designs



Genome Analysis and Database Update system

Open Science Grid

- Request: 1000 CPUs for 1-2 weeks. Once a month.
- 3 different applications: BLAST, Blocks, Chisel.
- Currently ramping up on OSG and receiving 600 CPUs and 17,000 jobs a week.



Common Middleware provided through Virtual Data Toolkit

Domain science requirements.

Globus, Condor, EGEE etc

OSG stakeholders and middleware developer (joint) projects.

Test on “VO specific grid”

Integrate into VDT Release.
Deploy on OSG integration grid

Include in OSG release & deploy to OSG production.

OSG allows you to

S

U

R

F

The Grid

Focus on a

Secure

Usable

Reliable

Flexible

Distributed Facility

Security - Management, Operational and Technical Controls

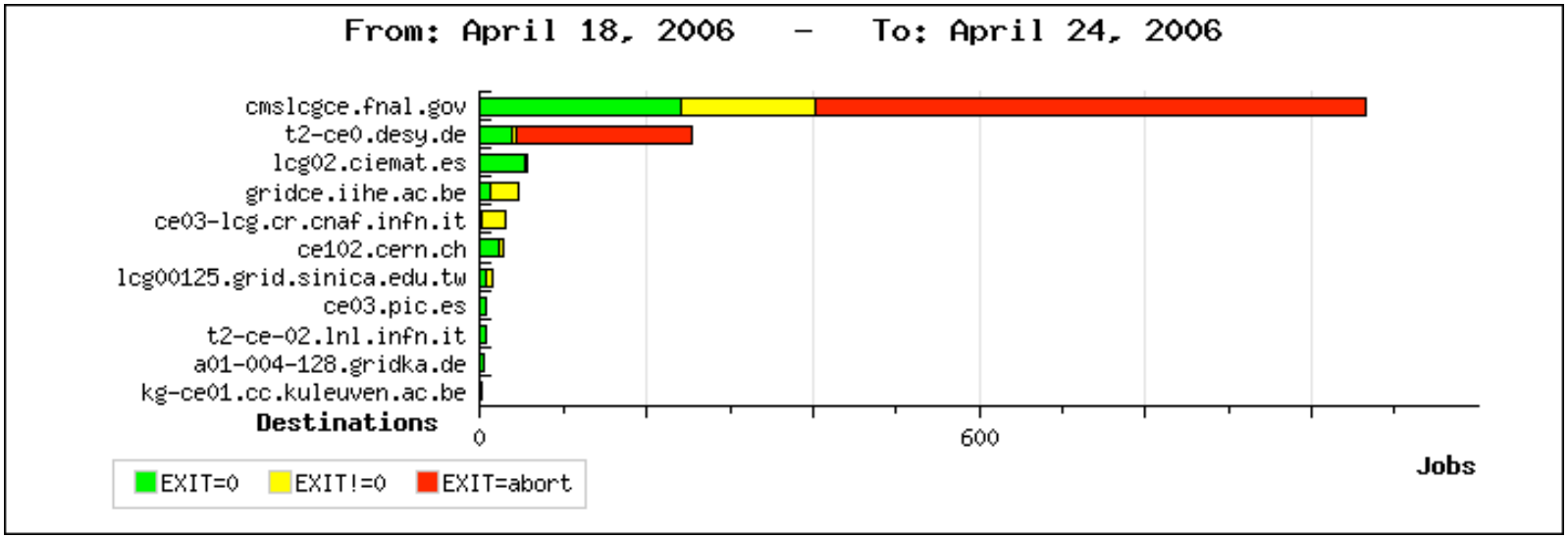
Open Science Grid

- *Management:* Risk assessment and security planning, Service auditing and checking.
- *Operational:* Incident response, Awareness and Training, Configuration management.
- *Technical:* Authentication and Revocation, Auditing and analysis. End to end *trust* in quality of code executed on remote CPU.

Usability - Throughput, Scaling, Fault Diagnosis

Open Science Grid

- Priority, quotas, write-authorization based on VO role rather than individual
- Minimize threshold of entry
- Address Errors through to the VO applications: e.g. CMS Analysis program dashboard:



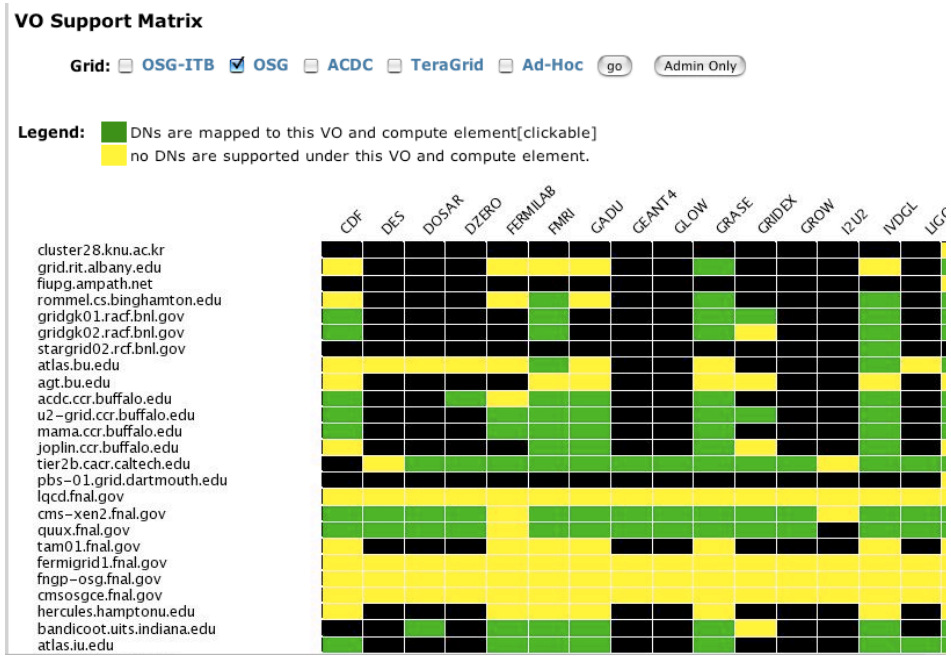
Reliability - Central Operations Activities

- Automated validation of basic services and site configuration Configuration of HeadNode and Storage to reduce errors:
  - Remove dependence on Shared File System
  - Condor-managed GRAM fork queue
- Scaling tests of WS-GRAM and GridFTP.

Daily Grid Exerciser:

The **Error Details** section gives more details on the errors encountered.

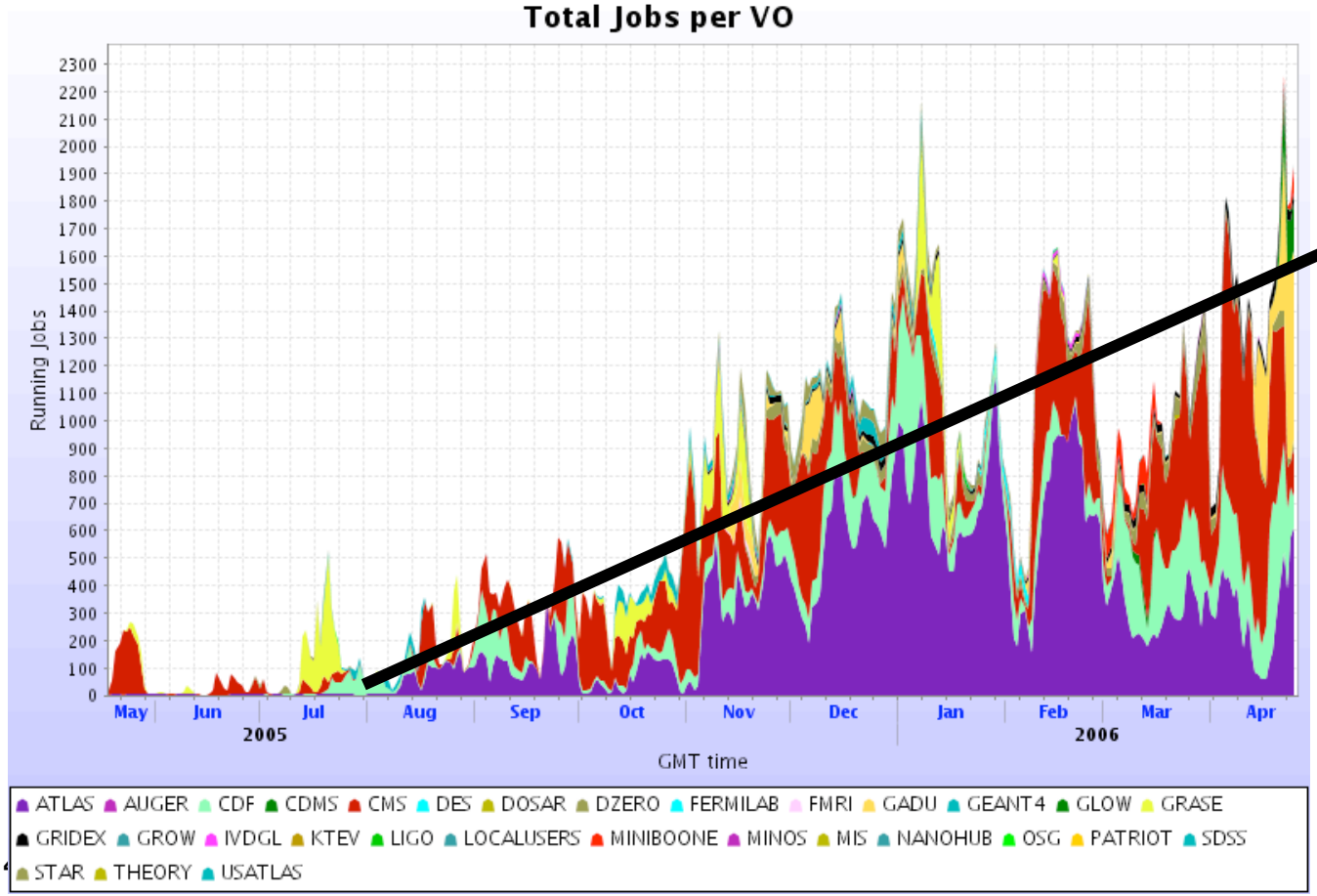
```
grid.dpcc.uta.edu
            590 Globus error 7: an authentication operation faile
           1748 Globus error 7: authentication with the remote se
```



The Future?

- Enable existing and future users.
- Increase the effective throughput.

First Year View:



Future: