Open Science Grid: Beyond the Honeymoon

Dane Skow
Fermilab

October 25, 2005

What is OSG ?

(Shortcut: EDG->EGEE = Grid3->OSG)

- OSG the Consortium
  - –Collaboration of contributors building the grid
- OSG the Grid
  - –The collection of services which do work
- OSG the collection of Virtual Organizations (VO)
  - –The scientist and the resources they control
- OSG the “project”
  - –The construction and persistent operation of the core

http://www.opensciencegrid.org

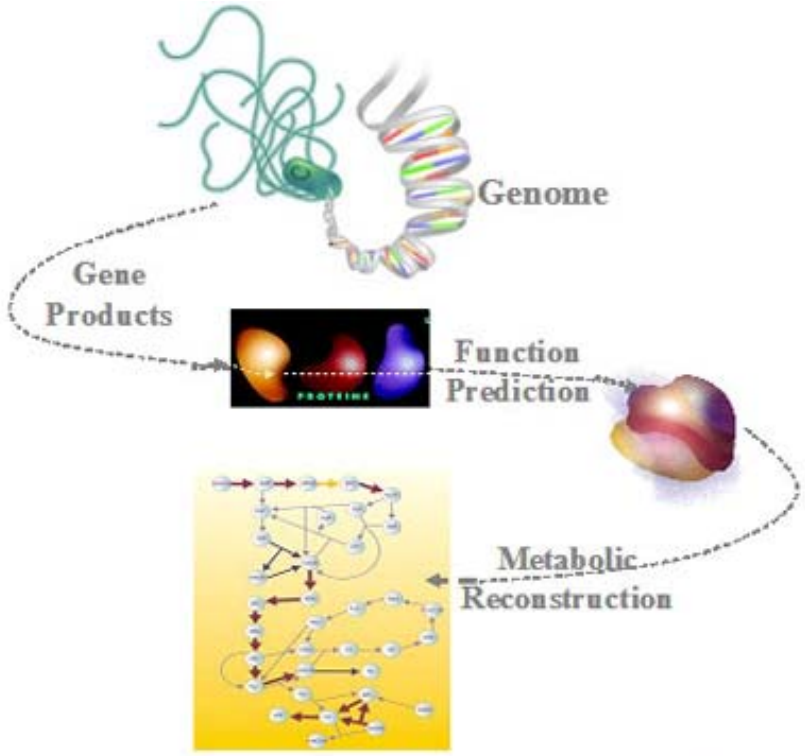
What's Happening ?

- July 20, 2005 "Grand Opening"
  - Currently operating a Grid
- Preparing Release 0.4 for December 2005
  - "Release" is a definition of a function set
- Integrating new participants
  - Council members: TACC
  - Grid participants: DOSAR, Dartmouth, Dark Energy Survey, Accelerator, Nanohub, …
- Developing Partnership plans
  - EGEE/LCG and TeraGrid primary focus

Who is using OSG ?

- The Virtual Organizations
  - High Energy and Nuclear Physics
    - CMS, ATLAS, STAR, DZero, CDF, Fermilab
  - Physics and Astronomy
    - LIGO, SDSS, Auger, DES
  - Biology
    - fMRI, GADU, GRASE, GLOW
  - Engineering
    - GRASE, GLOW
  - Computer Science
    - iVDGL, GLOW
- User Support is entirely provided by the VOs

One science success story



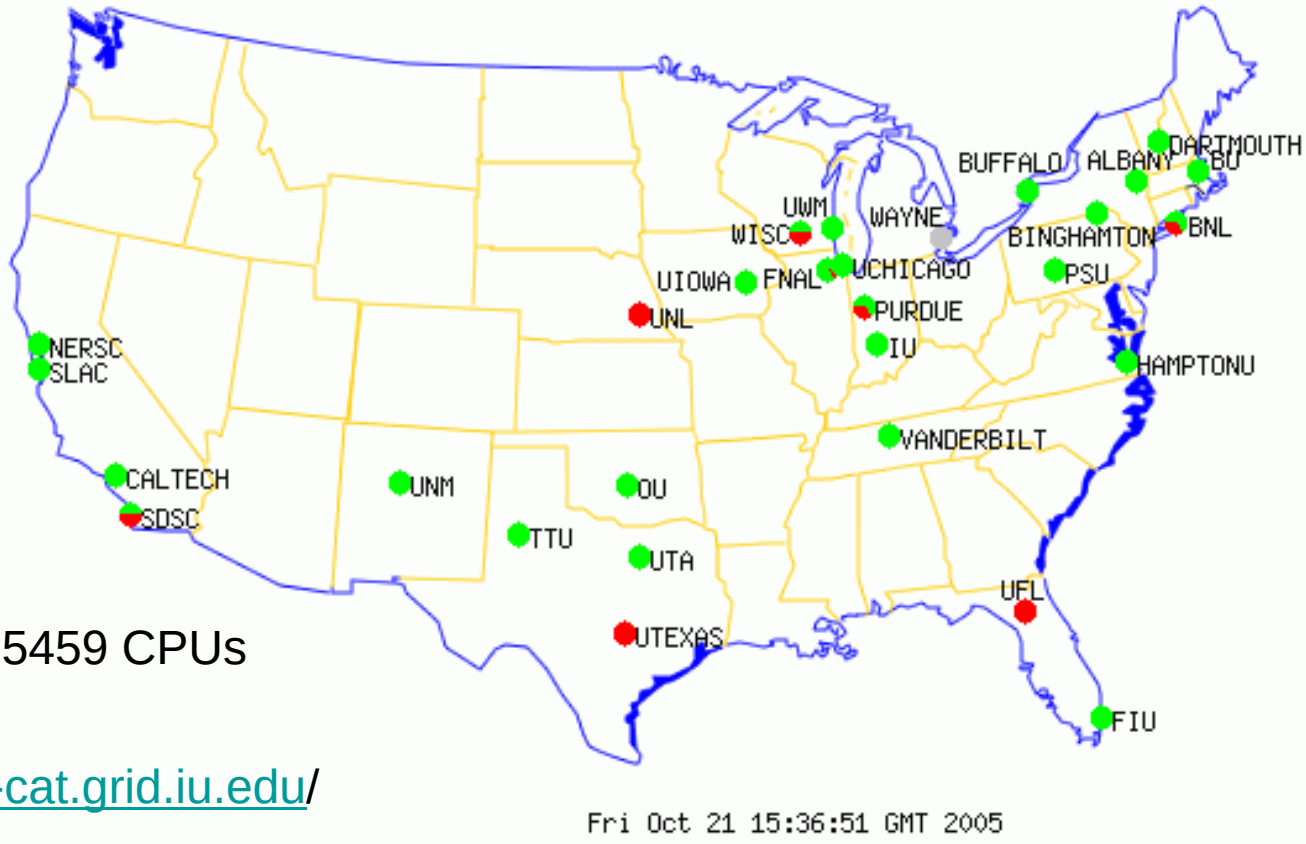
High-Throughput Genome Analysis using GADU/GNARE

- Genome Analysis and Update tool (GADU)
  - Comparative analysis of newly published genomes against known set
  - 1 bacteria genome sparks 12,000 step processing job
  - Auto processing frees researchers
  - Rapid turn around and publication boosts whole community
- See article in Science Grid This Week: http://www.interactions.org/sgtw/2005/0713/iow_20050713.html

Learning to Share

- Share the resources
  - Expressing policy clearly is tough
  - Biggest concern is access at time of owner need
- Share the information
  - Willingness, but different terminology and assumed context
  - Have to avoid “lecture mode”
- Share the frustrations
  - Have to avoid the “blame game”

OSG Production

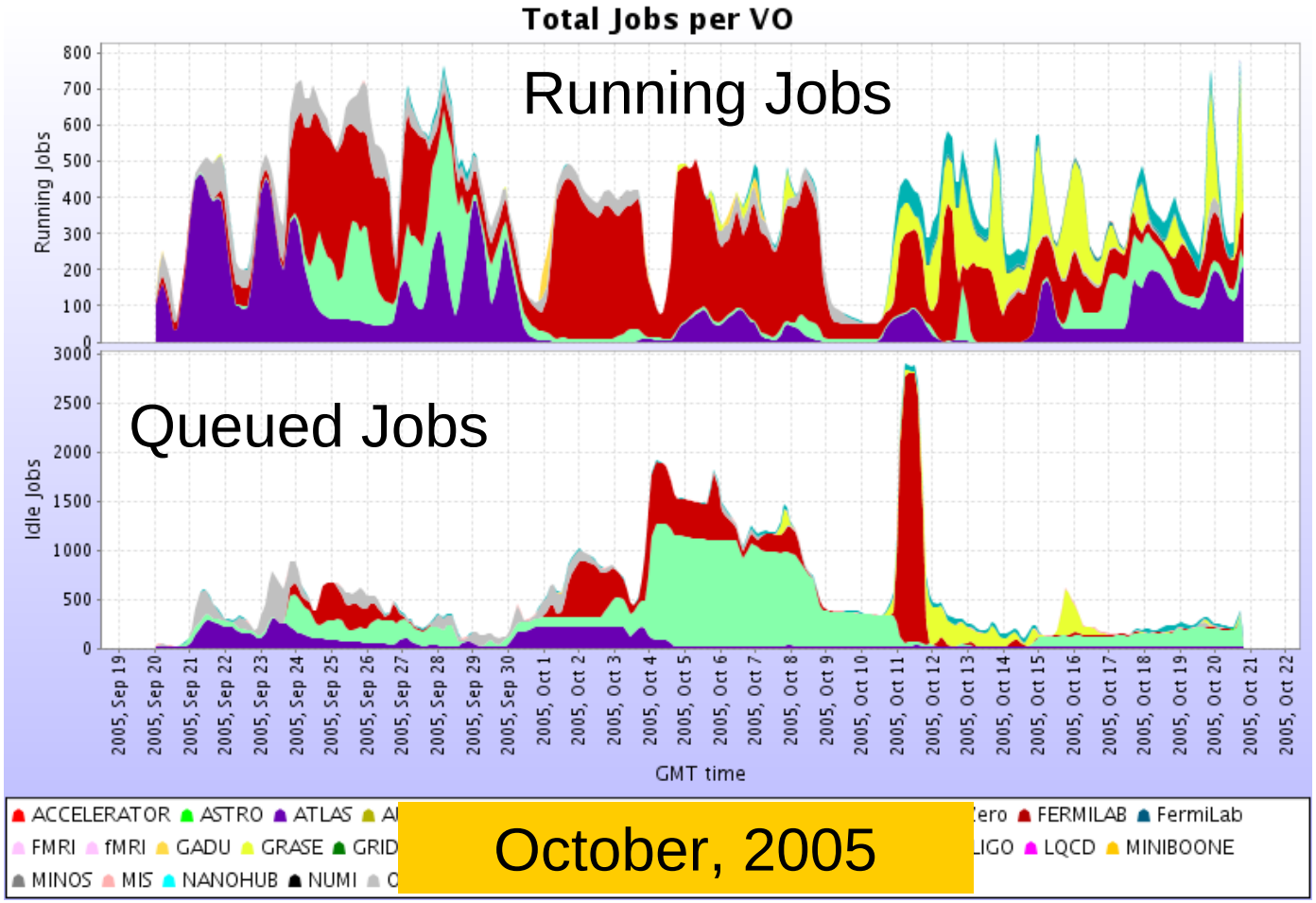


46 CEs, 15459 CPUs

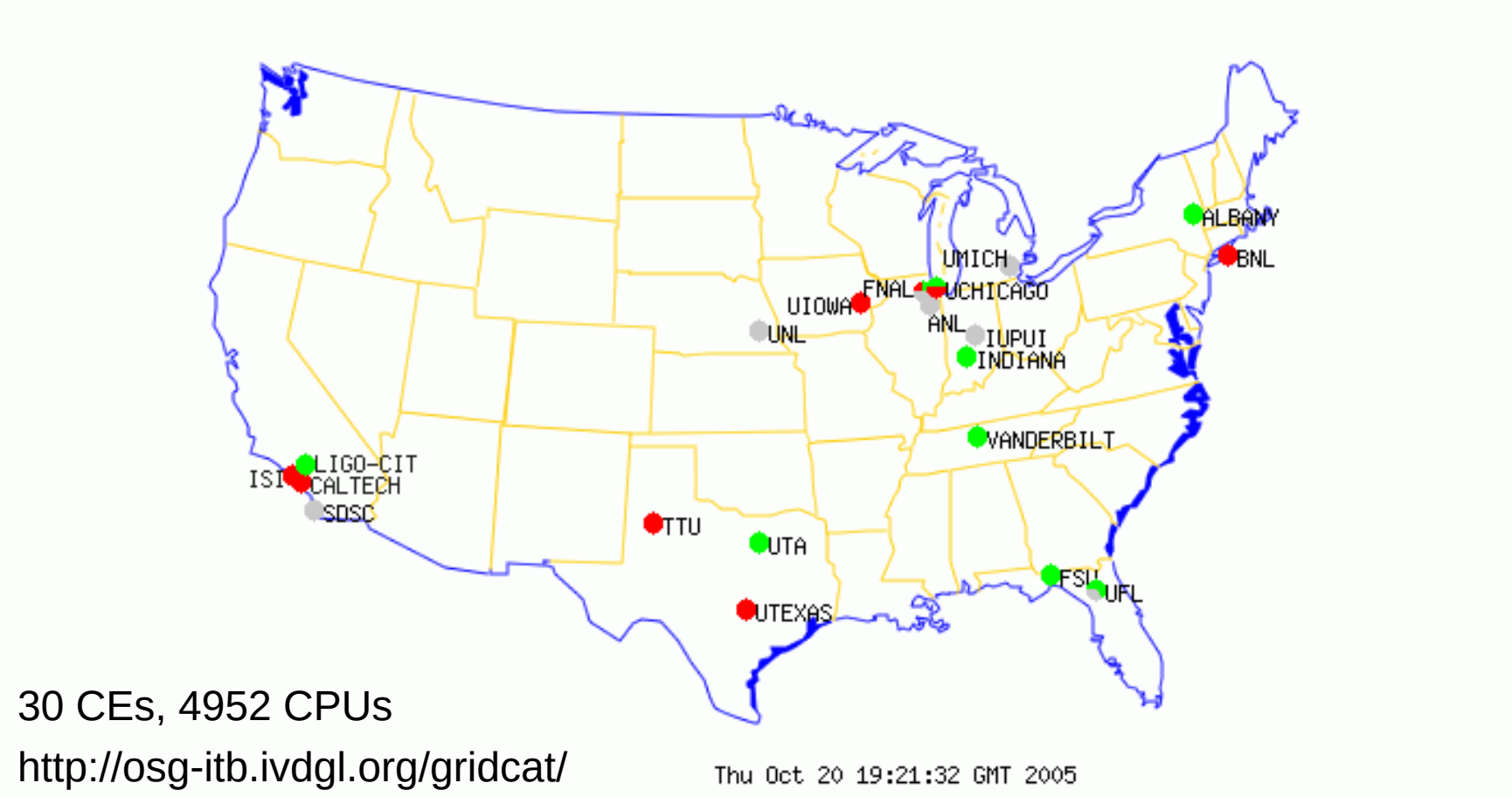
6 SEs

http://osg-cat.grid.iu.edu/

Recent Production Statistics



OSG Integration



30 CEs, 4952 CPUs

http://osg-itb.ivdgl.org/gridcat/

Service Oriented Architecture

"Interfaces not Implementations"

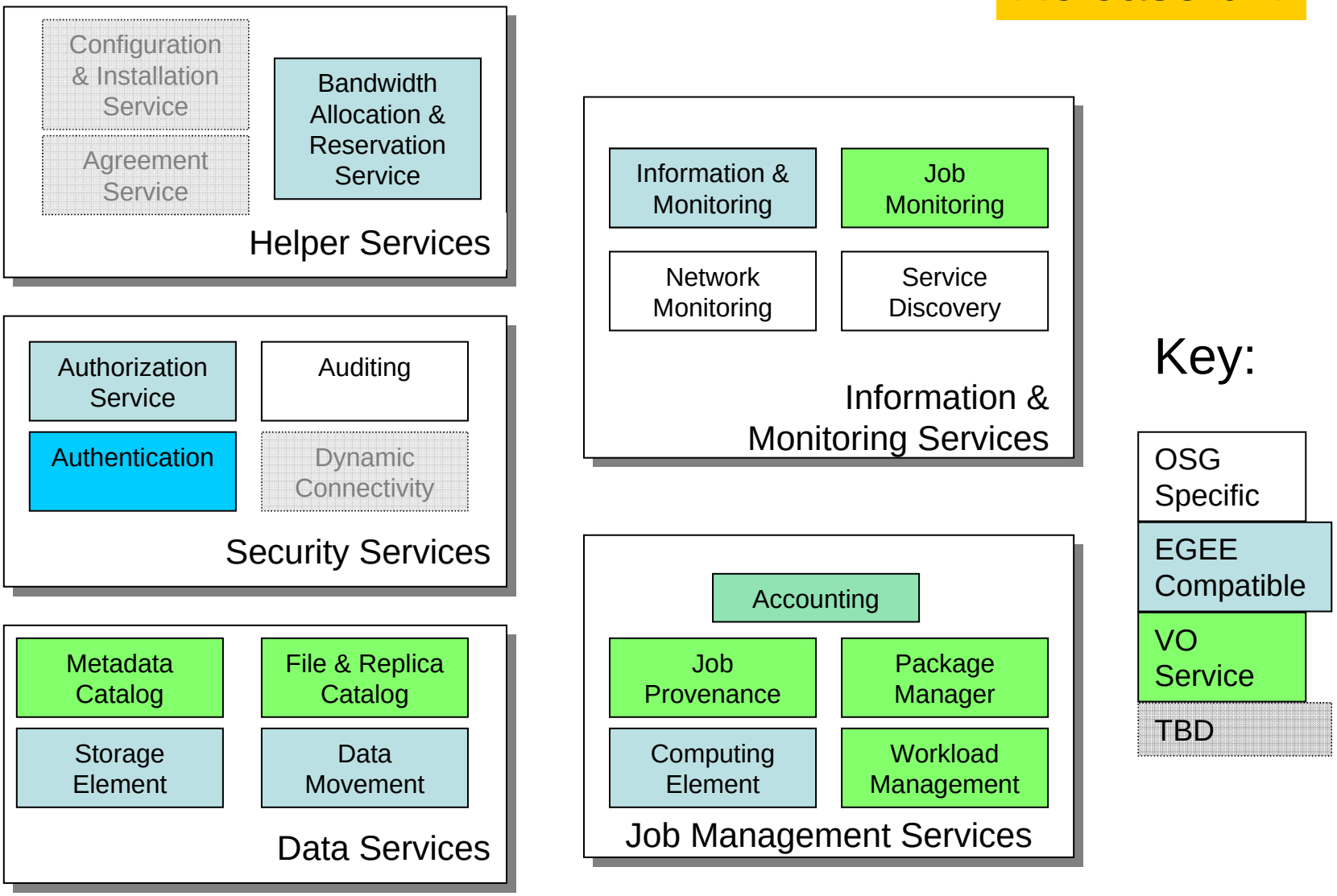
- We strive, with mixed success, to define functional requirements to allow multiple implementations
- Pragmatic pressures often require specification by implementation
- We are strongly guided by the EGEE architecture
- Path to Nirvana likely to be long…

Standards-based, open software

- However, inherent tension:
  - Evolving community means new ways of doing things
  - Production requirement value stability highly
- OSG Software
  - VDT is primary software integrator
  - Pacman used for software distribution
  - Software distributed is reference implementation
  - Integration Testbed (ITB) highly valued
- OSG infrastructure does not distribute nor manage application software

OSG Services

Release 0.4



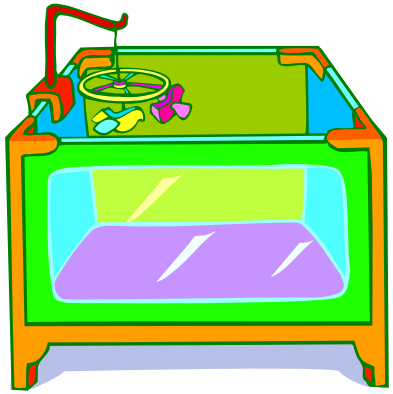
Release 0.4 Focus Areas

- Deployment Document (OSG Doc #300) is definition document of functional requirements for services

- Storage
  - CE/SE Data Integration
  - Access Methods
  - Authorization and Access Control
  - Resource Allocation Management
  - High throughput network utilization

Release 0.4 Focus Area

- Edge Services Framework
  - VO's need specialized persistent services to most efficiently utilize resources (CE)
  - Provide framework for providing VO control of a sandboxed environment
    - Working with Globus team on Xen Virtual Machine technology
    - First focus on operations advantages, then on security options



Release 0.4 Focus Area

- Web Services
  - VDT 1.3.7
    - incorporates Globus v4
      - WS GRAM
      - MDS4, service container, …
    - Deploying now on ITB
  - Simplify integration and evaluation of gLite components
  - Evaluate and understand Clarens
  - Begin understanding and evaluation of OGSA

Release 0.4 Focus Area

- Operational Efficiencies
  - Grid function
    - Higher utilization
    - Performance metrics
      - not robust
      - nor well understood
  - OSG growth
    - Current operations overhead too high
      - Registration human mediated
      - Monitoring and response as well

Release 0.4 Focus Area

- Workload Management
  - Service Catalogs
    - Clarens Discovery service
    - GridCat
  - Job Placement
    - Most everyone using Condor tools
  - Accounting
    - Have agreed on GGF UR format with EGEE and TeraGrid

- Remains VO scope component in this release

What's Next ?

- Develop sustaining funding plan for core infrastructure
  - Have no core staff now.
  - Meeting with US funding agencies this week
- Develop "Grid of grids" framework
  - EGEE model currently direct contact of resources
  - TeraGrid working on gateway with OSG
- Determine participation thresholds/model
  - TANSTAAFL ("There Ain't No Such Thing As A Free Lunch" - Robert Heinlein)
- Answer the "What's in it for me?" question
  - How do we broker exchanges of value ?
- Oh yes, and get the science done !

How will this Grid be Used ?