# Challenges for Grids

Markus Schulz
CERN IT GD
LCG/EGEE

# Disclaimer

- All views expressed are mine and are not necessarily shared by the projects or organization that I am associated with
  - Don’t blame: EGEE, LCG, CERN….
  - Critique, flames, and the like should be directed to:
    - Markus.schulz@cern.ch

# Approach

- Thinking a few years ahead
  - Based on what we know
  - Ignoring problems like
    - software quality (far from perfect)
    - lack of fabric management on sites
    - site admin fear of loosing total control
  - Focused on structural problems
    - Make production grids work at the required scale
    - Expand the systems to other domains
      - Industry, micro Vos, ……
    - Move closer to the grid vision

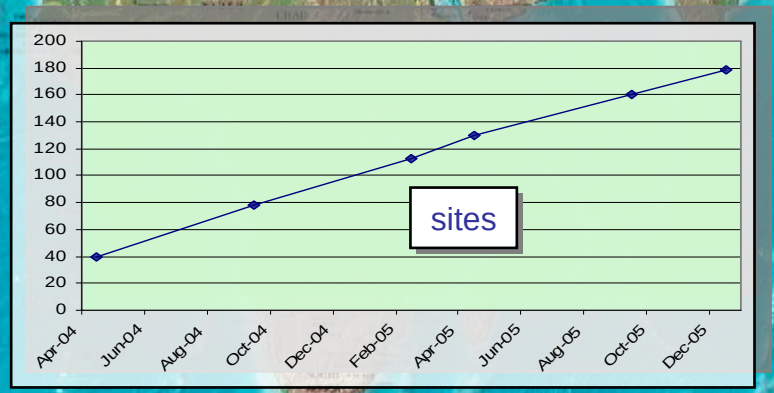
# Babylonian Confusion

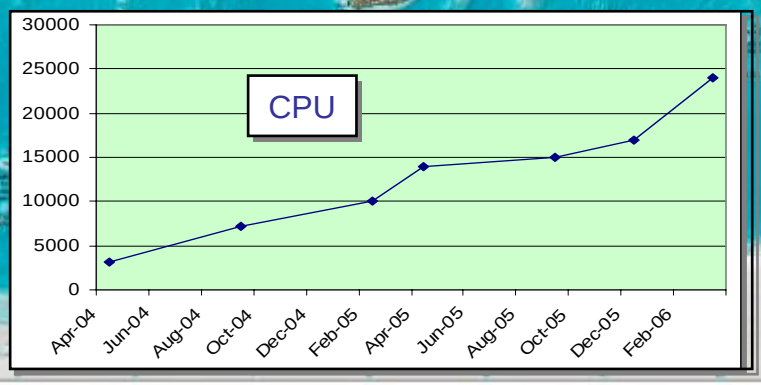
- What is called Grid covers☐:
  - Standalone Clusters
  - Clusters for scaling a single service
  - Intra organizational clusters
    - With central administrative control
  - Community computing
    - SETI@home, boinc
  - I.Foster: <------- This is what I will use…..
    - “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. “
    - ”On-demand, ubiquitous access to computing, data, and services”

# The Dangers of Success

- Early Success
  - Constraints from existing infrastructures
    - Users depend on them
  - Research ---> Production transition is very hard
  - Restricts standardization
    - The curse of backwards compatibility
- Example EGEE, WLCG, OSG, ARC
  - > 70 VOs

# EGEE Grid Sites : Q1 2006




**EGEE:**

> 190 sites, 40 countries

> 24,000 processors,

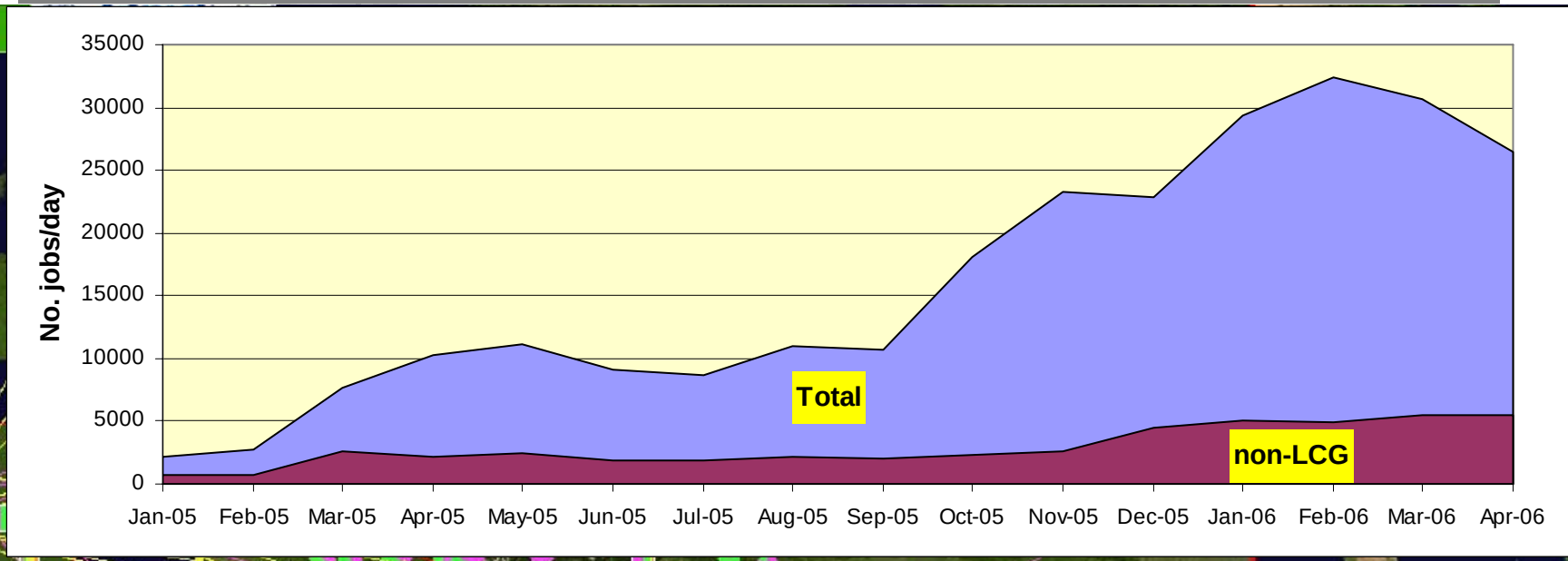
~ 5 PB storage

~ 70 Virtual organizations

# EGEE Operations

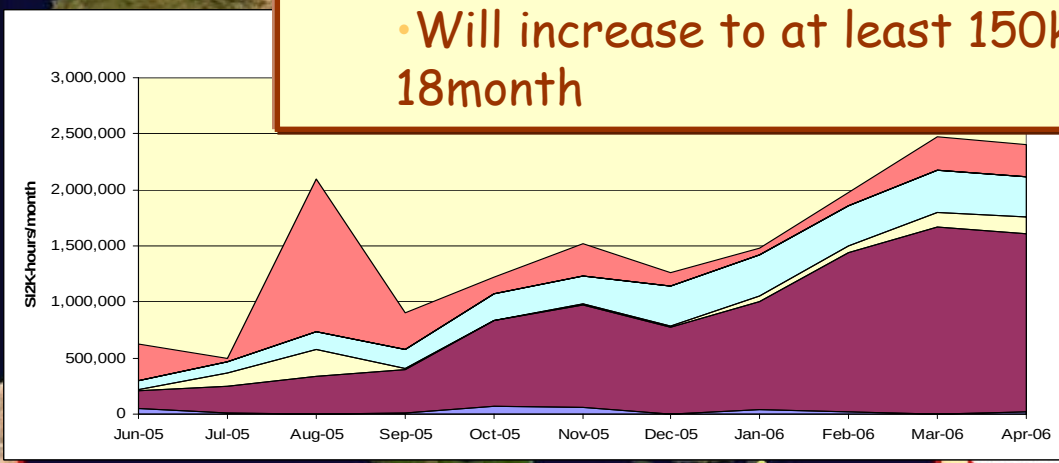
- Grid operator on duty
  - 6 teams working in weekly rotation
    - CERN, IN2P3, INFN, UK/I, Ru,Taipei
  - Crucial in improving site stability and management
  - Expanding to all ROCs in EGEE-II
- Operations coordination
  - Weekly operations meetings
  - Regular ROC managers meetings
  - Series of EGEE Operations Workshops
    - Nov 04, May 05, Sep 05, June 06
- Geographically distributed responsibility for operations:
  - There is no “central” operation
  - Tools are developed/hosted at different sites:
    - GOC DB (RAL), SFT (CERN), GStat (Taipei), CIC Portal (Lyon)
- Procedures described in Operations Manual
  - Introducing new sites
  - Site downtime scheduling
  - Suspending a site
  - Escalation procedures
  - etc

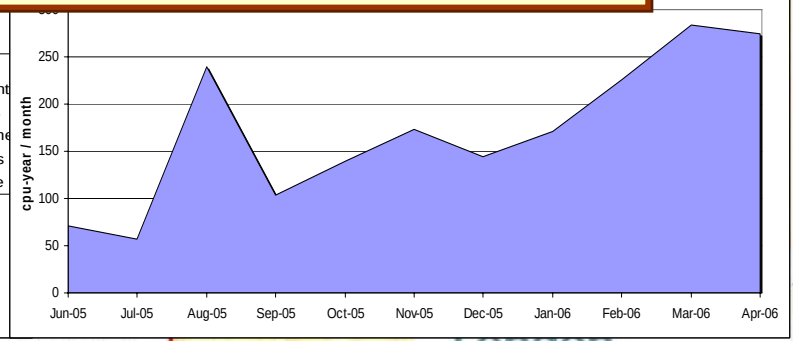
# Use of the infrastructure




Sustained & regular workloads of >30K jobs/day

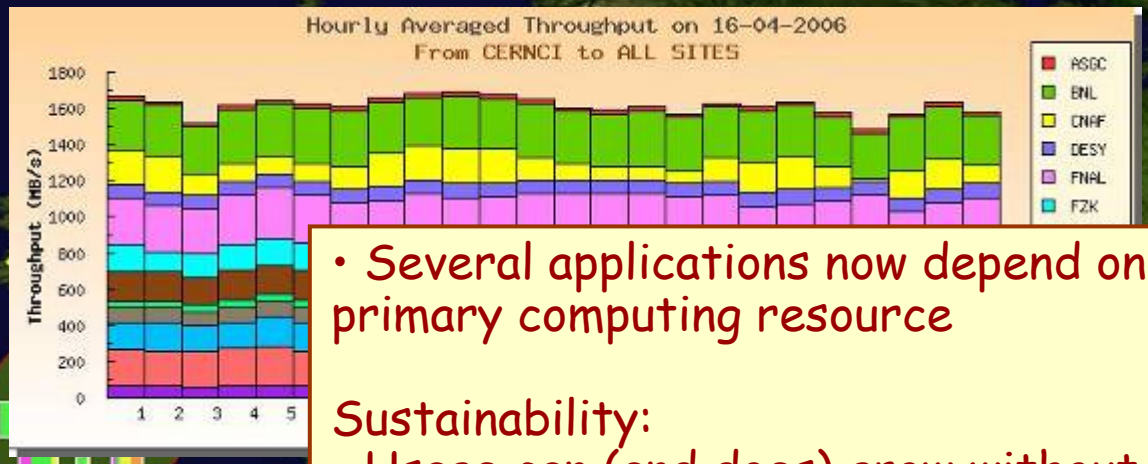
- spread across full infrastructure
- doubling/tripling in last 6 months – no effect on operations
- Will increase to at least 150k jobs/day in the next 18month



# Use of the infrastructure

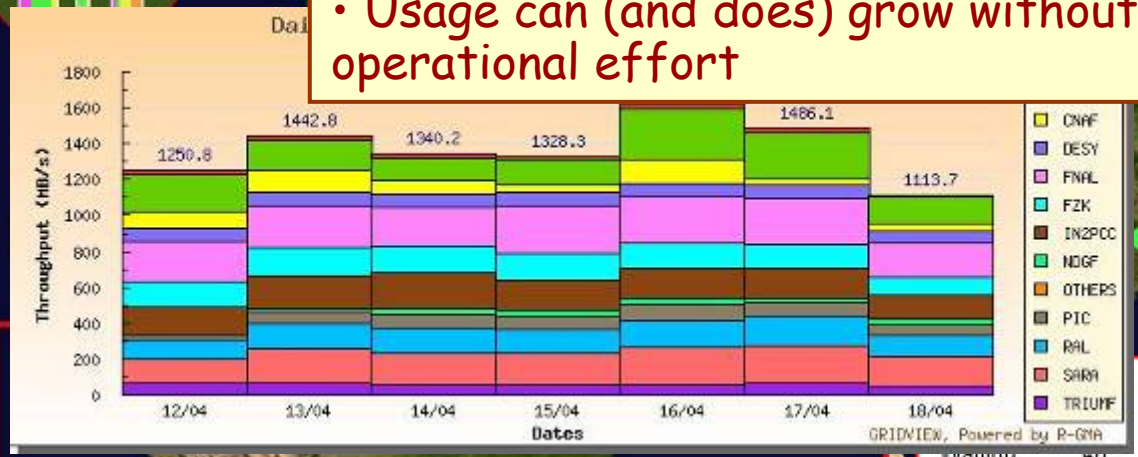
Massive data transfers > 1.5 GB/s




- Several applications now depend on EGEE as their primary computing resource

Sustainability:

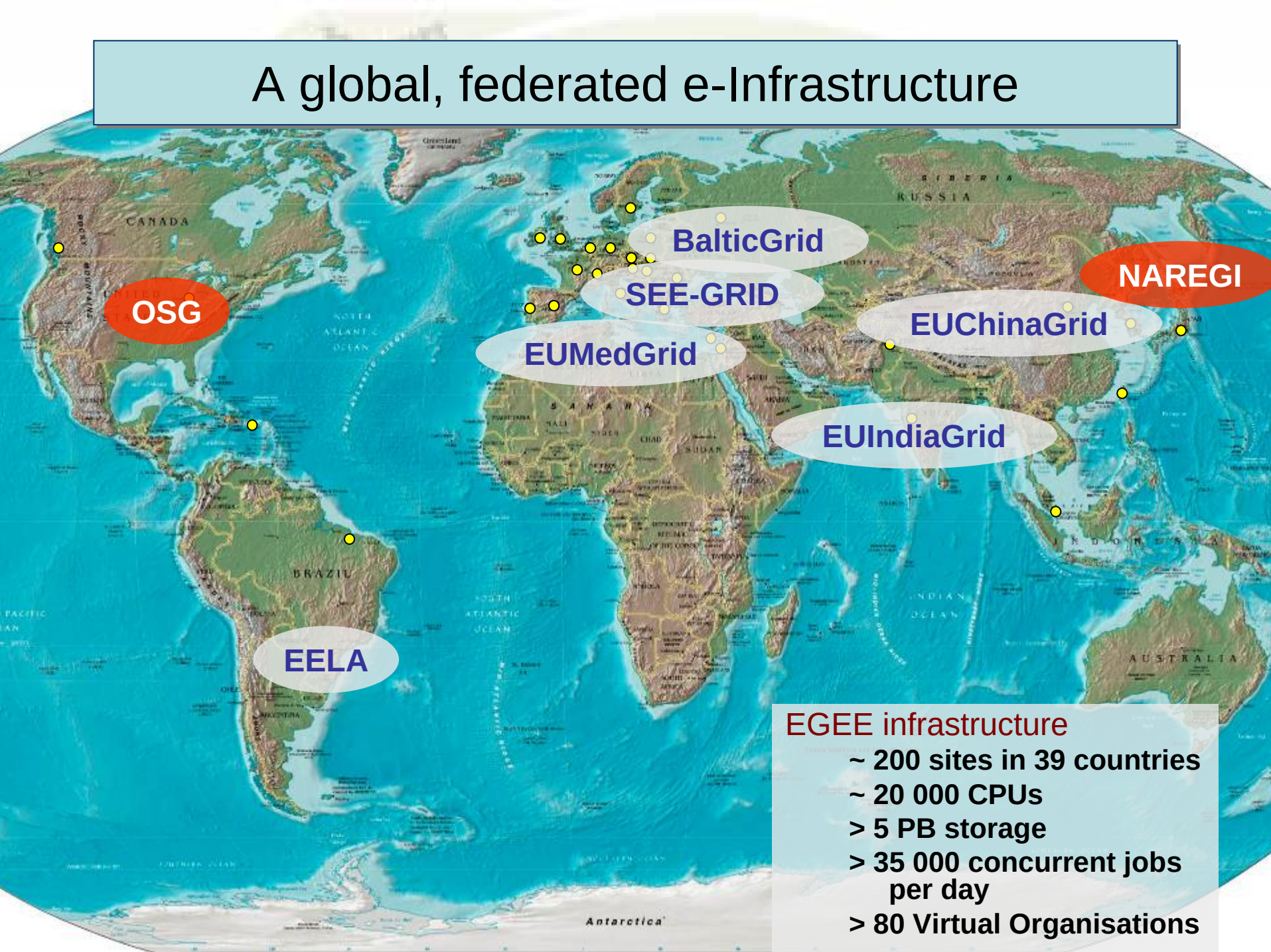
- Usage can (and does) grow without need for additional operational effort



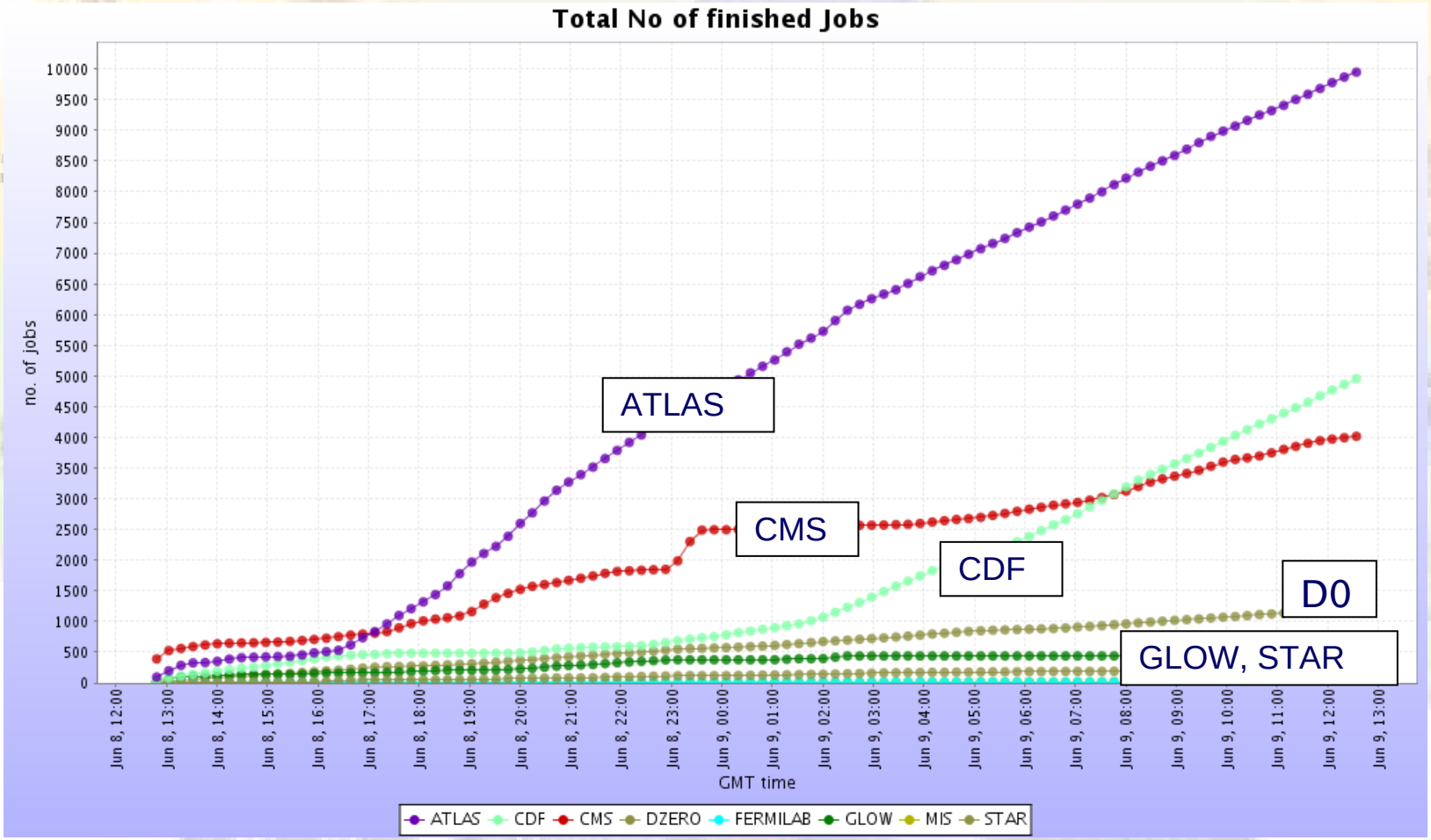

| | |
|---|---|
| Ready: | 750 |
| Scheduled: | 8840 |
| Running: | 11804 |
| Done: | 9347 |
| Aborted: | 4526 |
| Cancelled: | 141 |
| Active Sites: | 144 : 36826 |

Developed by e-Science, HEP
Imperial College London

# A global, federated e-Infrastructure



# OSG- Currently ~20,000 Jobs/Day



# This all looks very promising….

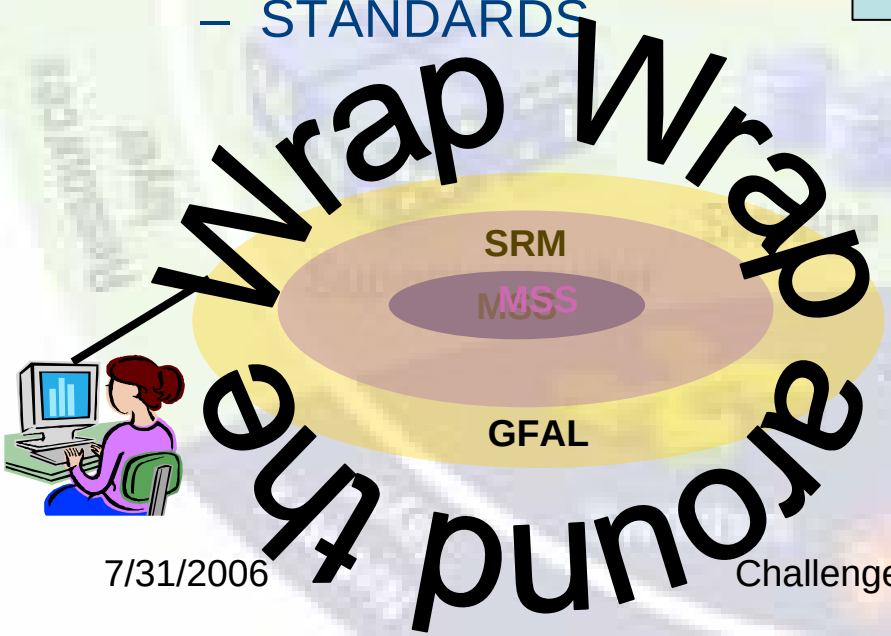
- But…….
  - Interoperation between grids
    - Lack of standardization
    - Several larger sites have to support multiple interfaces
  - Managing diversity inside grids
    - OS versions
      - Applications are sensitive and sites have preferences
      - Sites and user move independently
    - Batch systems
      - Each requires extensive work to interface
      - Limited to smallest set of shared functionality
        - » Frustrates users AND resource managers
        - » Lack of standardization

# More problems….

- Storage, DBs…
  - Different storage management systems are established
    - HSMs, disk pools with shared file systems
  - Different security, storage models, lack of standards
- VO management
  - Creation of a VO is straight forward
  - Getting access to resources requires:
    - Negotiation with resource providers
    - Significant effort of sites to host an additional VO
  - Accounting, dynamic prioritization, quotas problematic
    - on global level (between different Vos)
    - inter-VO
    - Constrained by national privacy laws
  - No market of resources

# More problems….

- Achievable reliability limited
  - The more complex services have to interact, the higher the probability that the overall service fails
- ‘Russian Doll Performance Sink’ here: File open
  - Applies to many services
- Grid interfaces need to be native interfaces
  - STANDARDS



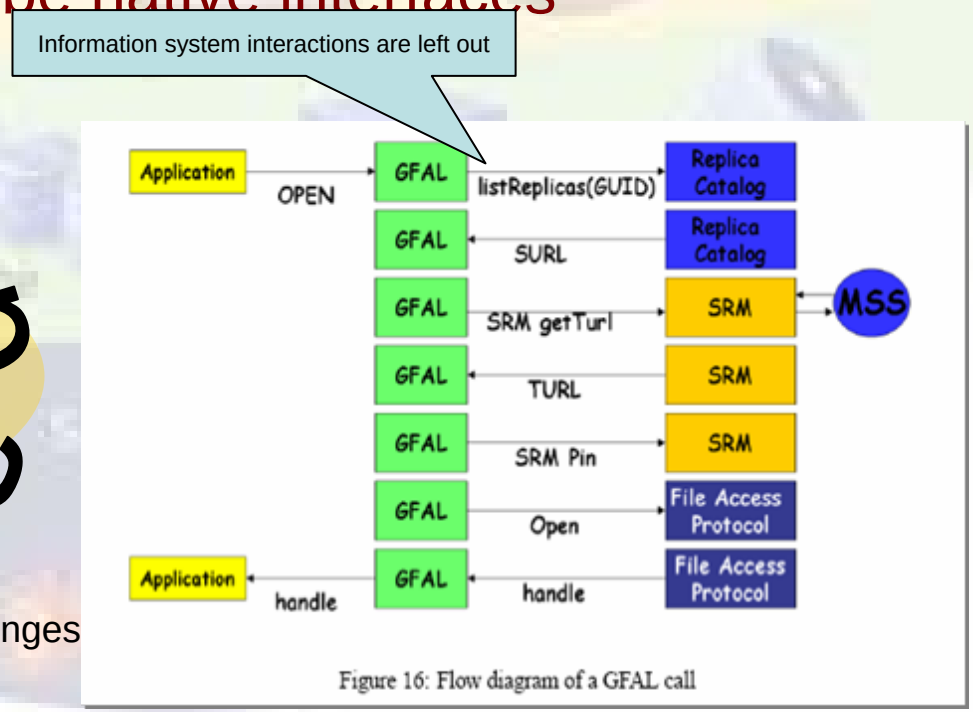



Figure 16: Flow diagram of a GFAL call

# State of Standardization

- First round of tentative standards
  - Mostly based on research work
    - Missed deployment and operations related part
  - Production grids started with 'de facto standards'
  - Now: OGSA
    - Much more detailed, recycles established standards
    - But: additional layers, old services will be wrapped!!!

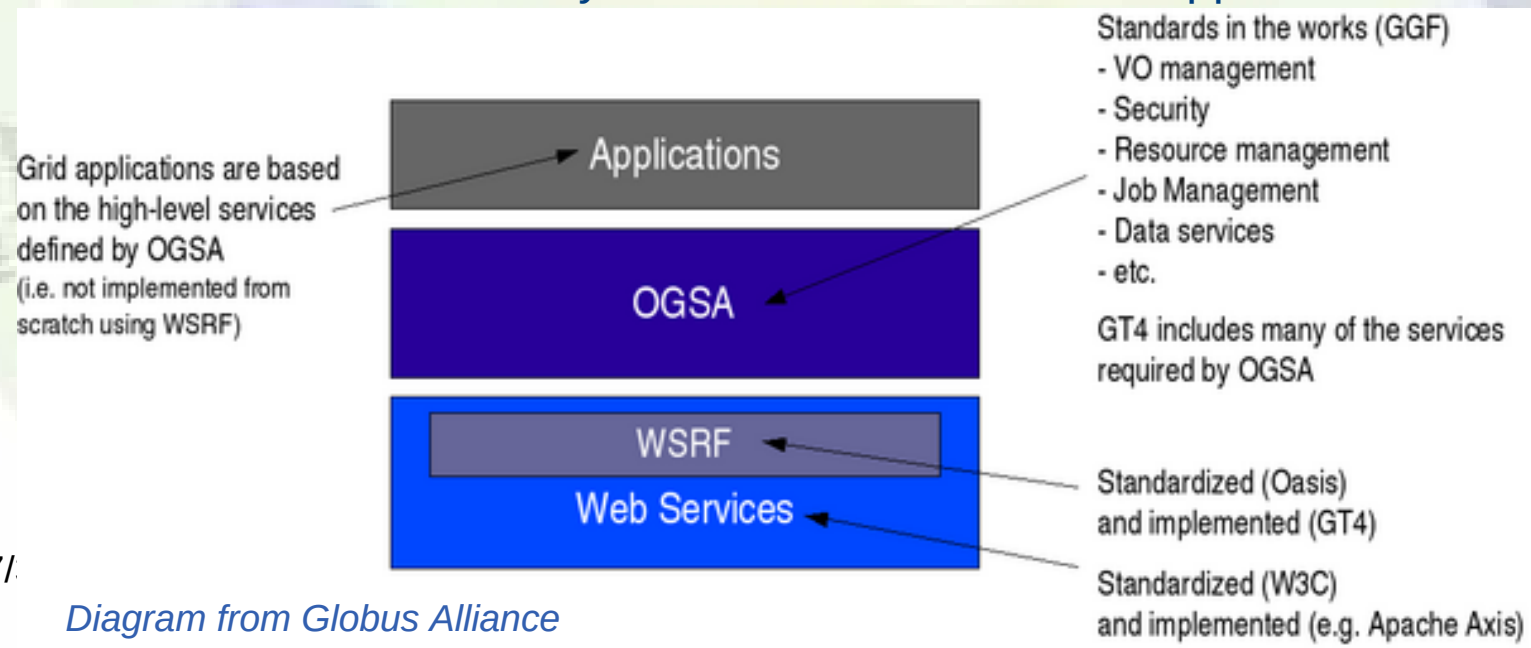
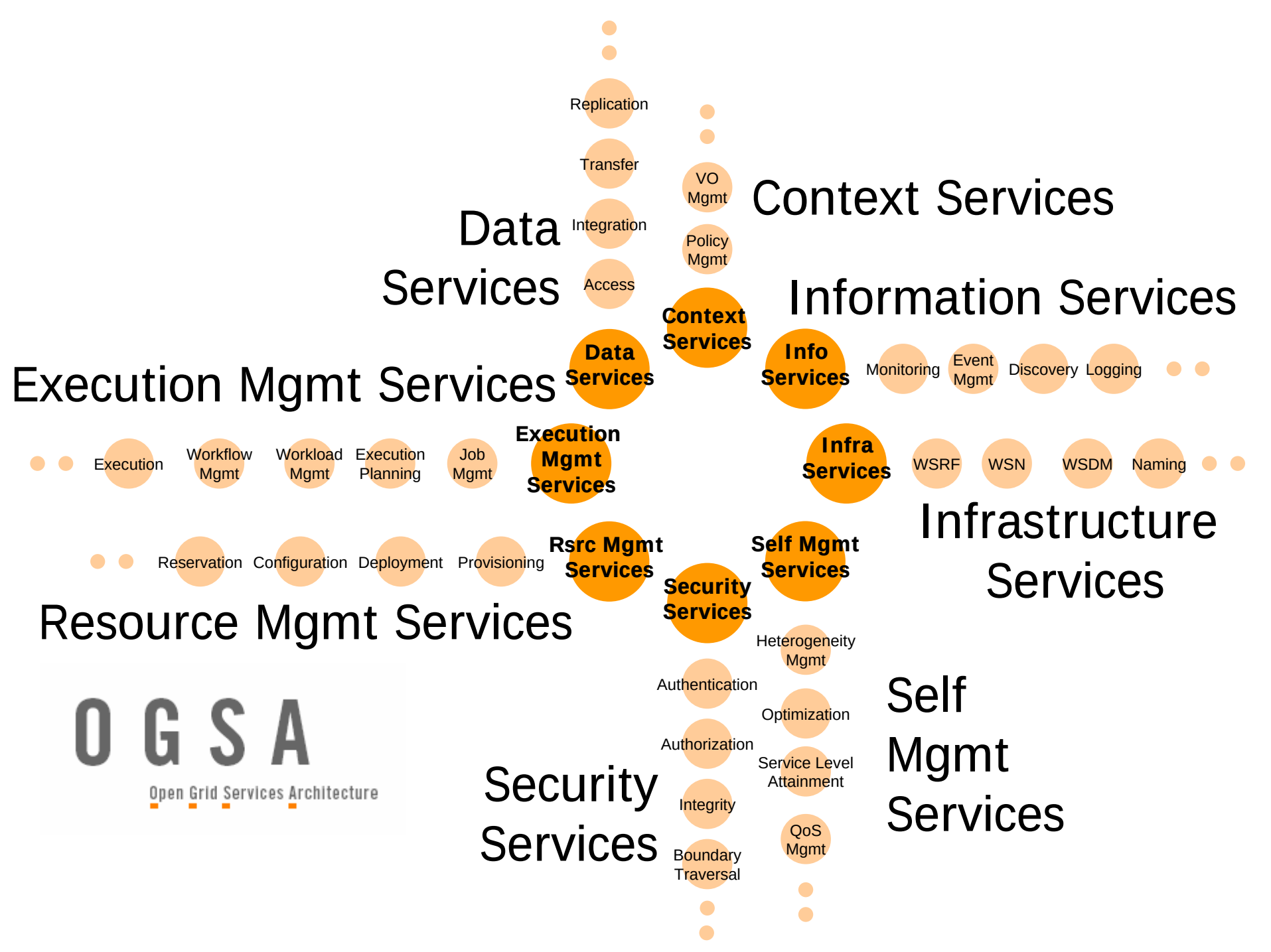
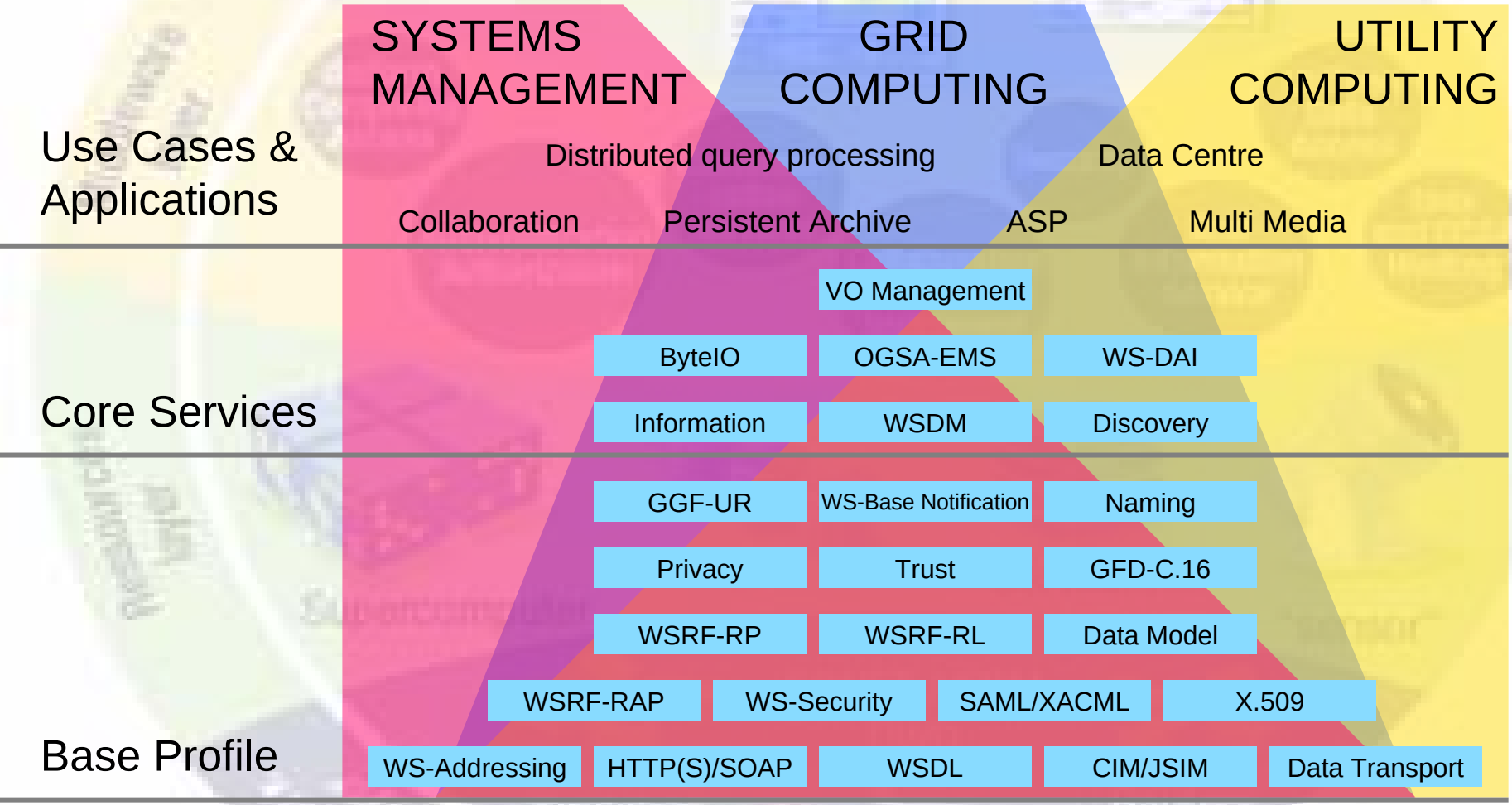



*Diagram from Globus Alliance*


Replication
Transfer
Integration
Access
Data Services
VO Mgmt
Policy Mgmt
Context Services
Context Services
Data Services
Info Services
Information Services
Monitoring
Event Mgmt
Discovery
Logging
Execution Mgmt Services
Execution
Workflow Mgmt
Workload Mgmt
Execution Planning
Job Mgmt
Execution Mgmt Services
Infra Services
WSRF
WSN
WSDM
Naming
Infrastructure Services
Reservation
Configuration
Deployment
Provisioning
Rsrc Mgmt Services
Self Mgmt Services
Security Services
Resource Mgmt Services
Heterogeneity Mgmt
Authentication
Optimization
Authorization
Service Level Attainment
Integrity
QoS Mgmt
Boundary Traversal
Security Services
Self Mgmt Services
OGSA
Open Grid Services Architecture

# Relevant Specifications




GRID Computing, Distributed Computing and Utility Computing are different views of the same **important** problem domain.

# Is there Hope?

- Diversity on OS level
  - Virtualization is making progress (XEN,…)
- Experience based standardization
  - Information systems,etc.
- Interoperation efforts start to influence standardization
- Core services start to work on native GRID interfaces
  - DBs, batch systems, storage
  - Still in an early state, but has a huge potential
    - Solid, well managed standards are needed
    - Otherwise a wrapper is the ‘best’ solution

# Detailed ‘Solvable’ Problem 1

- Easy introduction and destruction of VOs is at the core of the grid vision
- We can ease the config work, but access to resources is still based on negotiations
  - N*M problem
- For VOs and resource providers a system is needed for:
  - Trading resources (resource against resource or money)
  - Managing global priorities
  - Managing priorities between different groups inside a VO
  - And the same for quotas
  - Needed for: CPU, Storage, and Bandwidth
  - Has to be dynamic and leave control with the resource owners
  - For Oil and frozen orange juice the problem has been solved….

# Illustration from HEP

- The ATLAS VO that has ~20 □□□□research groups (b-Physics, top, higgs…)
  - The members of these groups have different roles (about 5)
    - User, storage admin, leading researcher…
- There are several experiments with similar structure
- The association can be expressed via the VOMS proxy extensions

- On Monday ATLAS has a standard split of:
  - 10% for b-Physics
  - 20% for top
  - 60% for Higgs
  - The rest equally split…
  - The lead researcher should get top priority
- Tuesday rumors spread that the student Judith from SUSY team of CMS has an indication of a signal (a signal is a ticket to Stockholm)
  - ATLAS needs now in almost real time:
    - Shift 90% of their resources and top priority to student Jack of their corresponding team
- Friday Judith gives a presentation in which she explained that she mixed the Monte Carlo Data with real data
  - ATLAS has to switch now quickly back to standard mode….

# The Resource Providers Story

- There are a few hundred or even thousands
- We pick one:
  - Computing center of the physics department of College Town
    - Funding by:
      - National grid project, departments budget which is in CMS, donation by the foundation for top-physics, …..
  - The center is open for all ATLAS and CMS groups
    - But, over a long time resources have to be provided based on funding
    - This is currently solved with static configuration of fair share schedulers
      - Because there is NO trading system or currency
    - The site can't change configuration on the fly
      - As most grid sites a fraction of an admin is running the grid aspect
- A system that would allow management of computing currencies and that would provide a market to establish a price would simplify the situation

# Detailed ‘Solvable’ Problem 2

- Access to storage
  - For large files, where latency is a minor issue solutions are underway
    - Interfaces to MSS, FTS for reliable transport, replica catalogue
    - Latency is on the order of several seconds to minutes
- Missing
  - The replacement for the users home directory on the grid
  - Characterization:
    - Many, many files ( > 10^6 per user)
    - Average size is small ( 1 MB per file, total from 1GB to a few 100GB)
    - In a work session the user will create several
    - And access quite a few O(100)
    - Access is almost ☐random
    - Latency matters since the user will work interactive with these files
      - Statistical data, plots, etc.
- Hint:
  - Central storage or replicating all files to all sites is not an acceptable solution