

Prelude

- Objectives
 - The purpose of this tutorial is to give a practical introduction to Grid Computing and the international EGEE grid.
 - The tutorial gives experience of the current middleware during practicals using the GILDA training testbed, as provided by INFN/University of Catania and partners.
- Approaches
 - The course has a practical emphasis, and is intended for anyone with an interest in grids, including researchers and systems administrators.
- Speakers:
 - Eric Yen - Eric.Yen@twgrid.org
 - Jinny Chien - Jinny324@gate.sinica.edu.tw
 - Hung-Che Jen - wfifang@gate.sinica.edu.tw
 - Min Tsai - mtsai@twgrid.org
- Organizers: <http://www.ngp.org.sg/gridasia/2007/main.php>
- Prerequisites:
 - Participants should have Linux user background and must bring their own laptops with an ssh client (e.g. PuTTY), a graphical browser (e.g. IE7) and a Java plug-in (1.5.x JRE) installed.

Agenda

Tuesday 05 June 2007




[top](#)↑

- <http://indico.cern.ch/conferenceDisplay.py?confId=15418>

08:30->09:00 Registration

09:00 Introduction to Grid & Overview of EGEE Project (30') Eric Yen

09:30 Middleware Overview (30') Min Tsai (ASGC)

10:00 Workload Management System (30') ( Slides  ) Hung-Che Jen (ASGC)

10:30 break

11:00 Authentication, Authorization and Security (30') Chen-Yi Chien





11:30 Using Certificate & Simple Job Submission (40') Chen-Yi Chien





log into User Interface:
glite-tutor.ct.infn.it
glite-tutor2.ct.infn.it

Accounts:
singapore01 ~ singapore40




JDL requirements:
requirement=(other.GlueCEUniqueID=="grid011f.cnaf.infn.it:2119/jobmanager-lcgpbs-short");
requirement=(other.GlueCEUniqueID=="iceage-ce-01.ct.infn.it:2119/jobmanager-lcgpbs-infinite");



12:10 Lunch

13:30 Information System (30') ( Slides   ;  practical) Min Tsai (ASGC)

14:00 Data Management (1h00') ( Slides   ;  practical) Hung-Che Jen (ASGC)

15:00 break

15:30 Put It Together (40') ( Slides  ;  practical) Min Tsai (ASGC)

16:10 Next Steps and Discussion (50') ( Slides ) Min Tsai (ASGC)

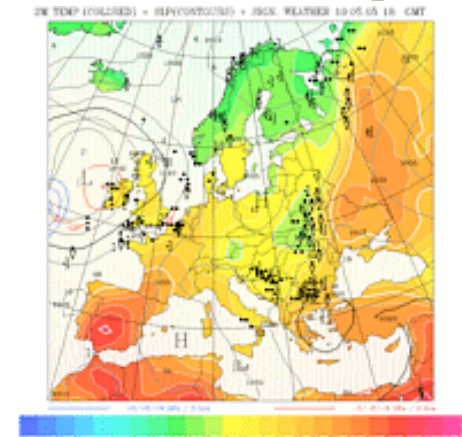
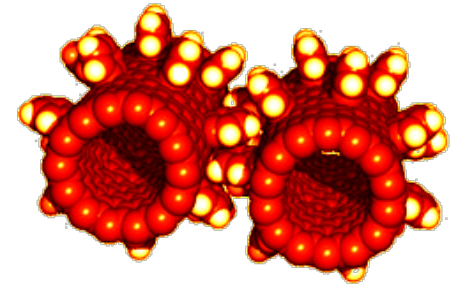
An Introduction to Grid Computing

Presented by Eric Yen

With thanks to Mike Mineter and EGEE colleagues for these slides

- **Introduction to**
 - e-Research and e-Science
 - Grid Computing
 - e-Infrastructure
- **Some examples**
- **Grid concepts**
- **Grids - Where are we now?**

- **Many vital challenges require community effort**
 - Fundamental properties of matter
 - Genomics
 - Climate change
 - Medical diagnostics
- **Research is increasingly digital, with increasing amounts of data**
- **Computation ever more demanding**
e.g.: experimental science uses ever more sophisticated sensors
 - Huge amounts of data
 - Serves user communities around the world
 - International collaborations



- **Collaborative research that is made possible by the sharing across the Internet of resources (data, instruments, computation, people’s expertise...)**
 - Crosses organisational boundaries
 - Often very compute intensive
 - Often very data intensive
 - Sometimes large-scale collaboration
- **Early examples were in science: “e-science”**
- **Relevance of “e-science technologies” to new user communities (social science, arts, humanities...) led to the term “e-research”**

People with shared goals

Improvised cooperation



People with shared goals

Email

File exchange

ssh access to run programs

Enabled by networks:

national, regional and

International: GEANT

**Collaborative
“virtual computing”**



Improvised cooperation



People with shared goals

**Sharing data, computers, software
Enabled by Grids – two main types**

- specific to a project
- supporting many collaborations

Email

File exchange

ssh access to run programs

Enabled by networks:

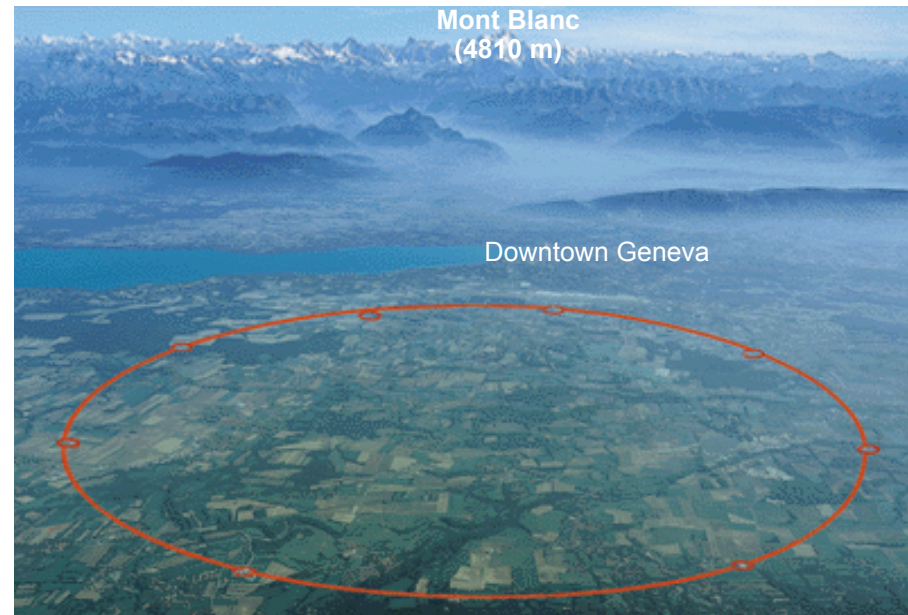
**national, regional and
International: GEANT**

- **Networks + Grids**
 - *Networks connect resources*
 - *Grids enable “virtual computing” - resource sharing across administrative domains*
 - *“admin. domain”: institute, country where resource is; system management processes;...*
- **+ Operations, Support, Training...**
- **+ Data centres, archives,...**

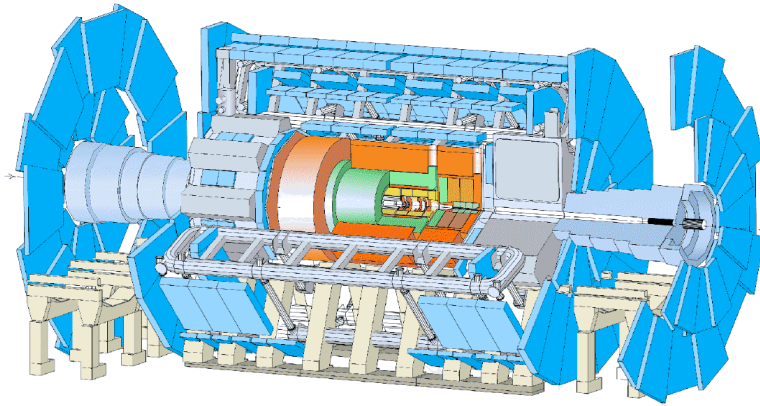
Some examples of e-science

- Large amount of data
- Large worldwide organized collaborations
- Computing and data management resources distributed world-wide owned and managed by many different entities

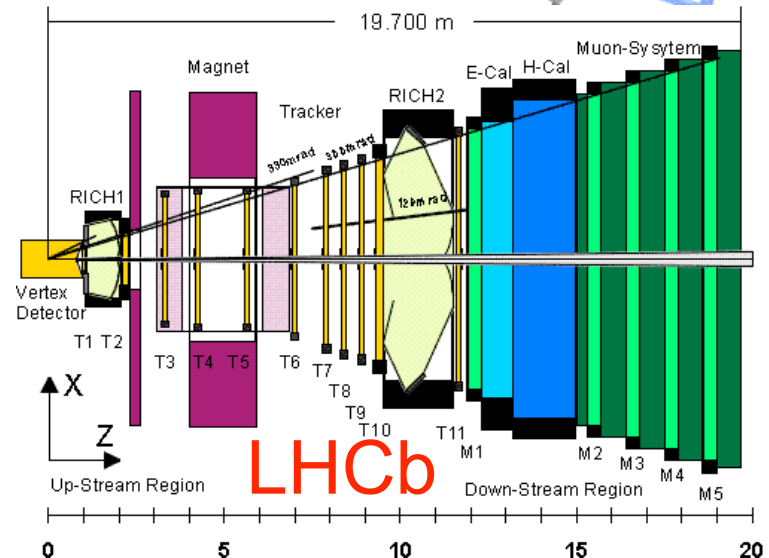
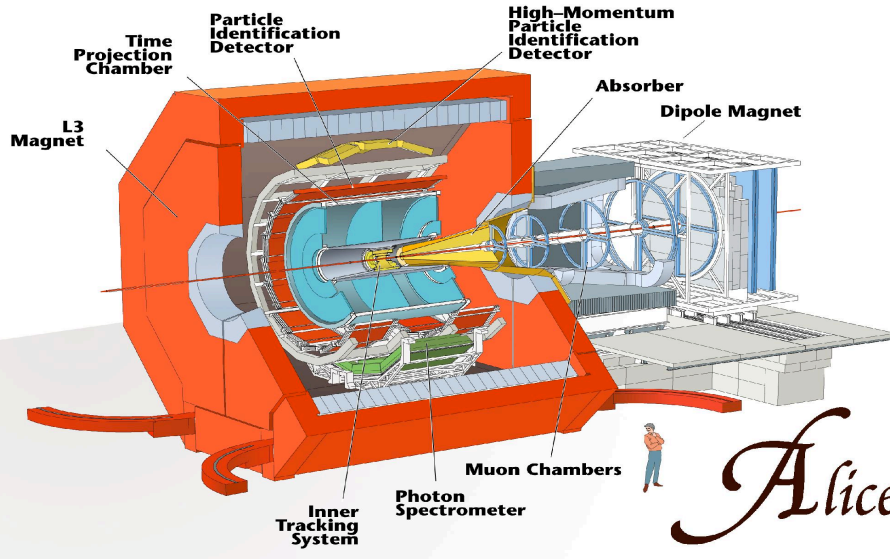
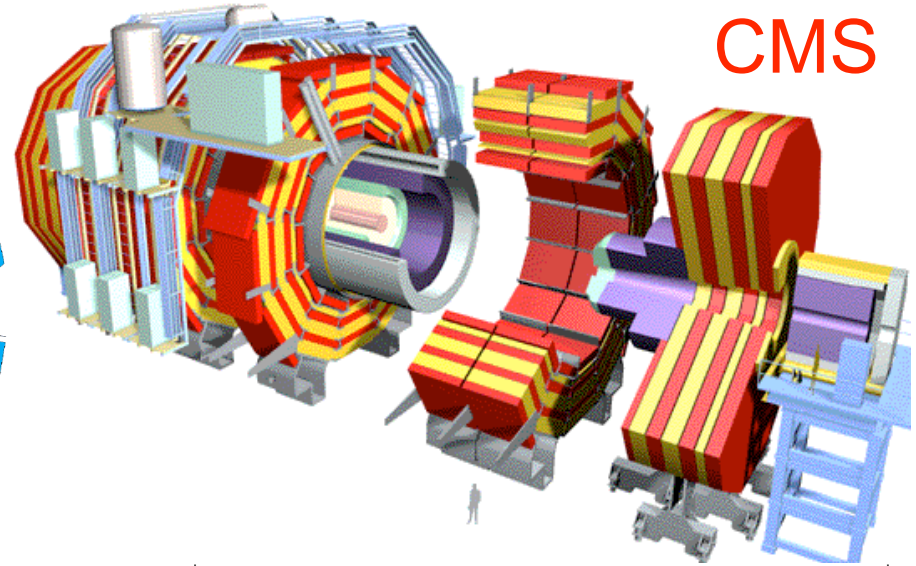
- Large Hadron Collider (LHC) at CERN in Geneva Switzerland:
 - One of the most powerful instruments ever built to investigate matter



ATLAS

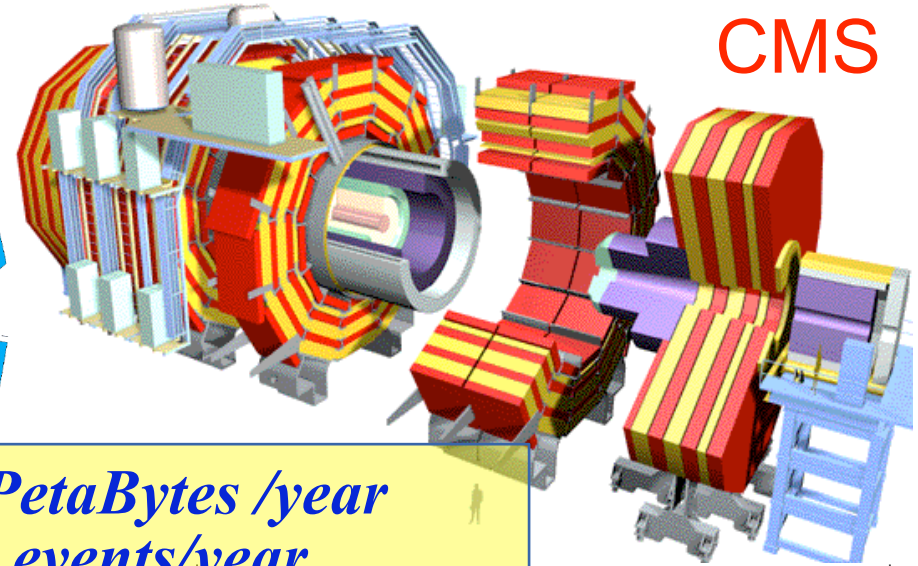
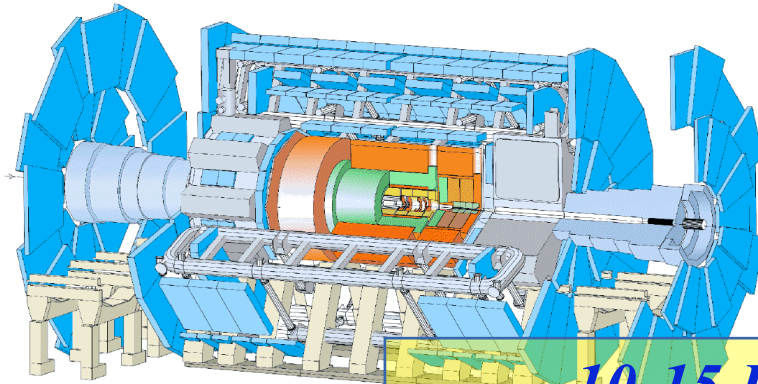


CMS

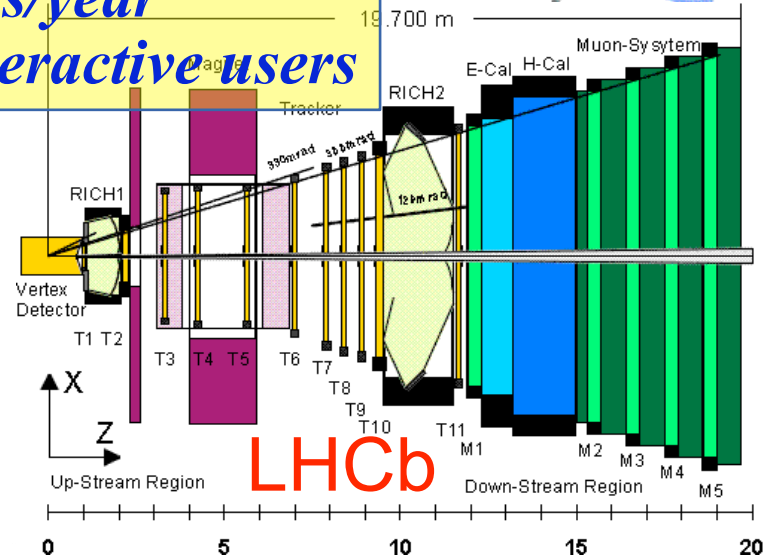
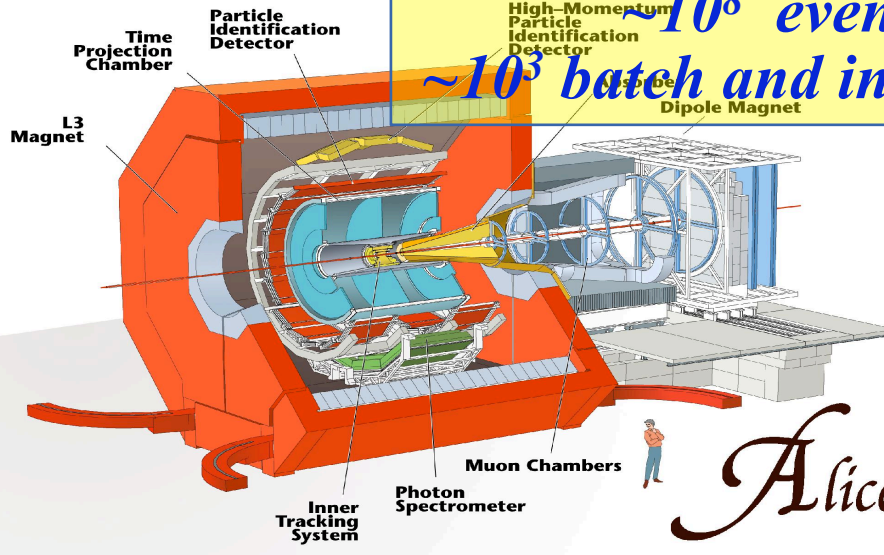


ATLAS

CMS



~10-15 PetaBytes /year
~10⁸ events/year
~10⁵ batch and interactive users



Alice

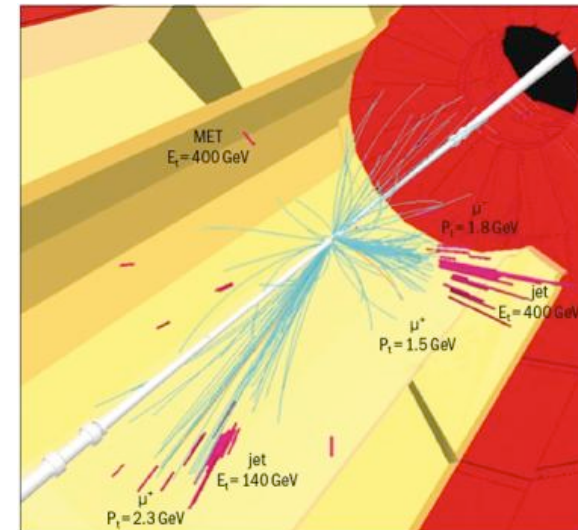
LHCb

- **Data management:**
 - Demonstrated data transfers at nominal rates: 1.6 GB/s through FTS
 - 1 GB/s with real (simulated) workloads
 - 2 large experiments transferred >1 PB/month in summer 2006

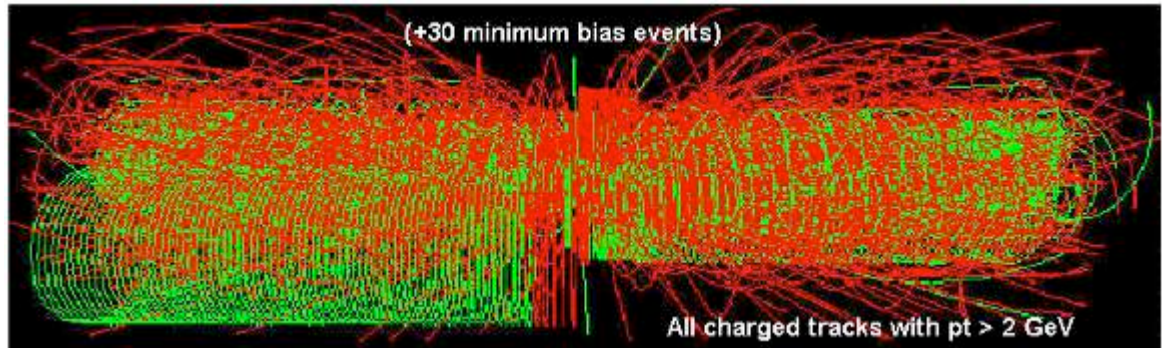
- **Workload management**
 - CMS – computing service challenge achieved 50k jobs/day
 - CMS aim this year for 100k jobs/day; ATLAS for 60k

- **Reliability and availability**
 - Significant effort to ensure Tier 1 sites meet MoU commitments – using site and service monitoring

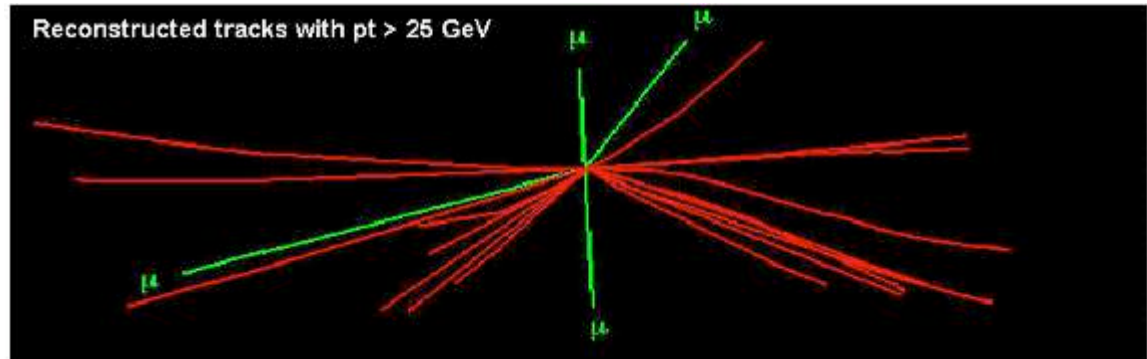
- **Grid is now the primary source of computing resources for LCG**



Starting from
this event



Looking for
this “signature”



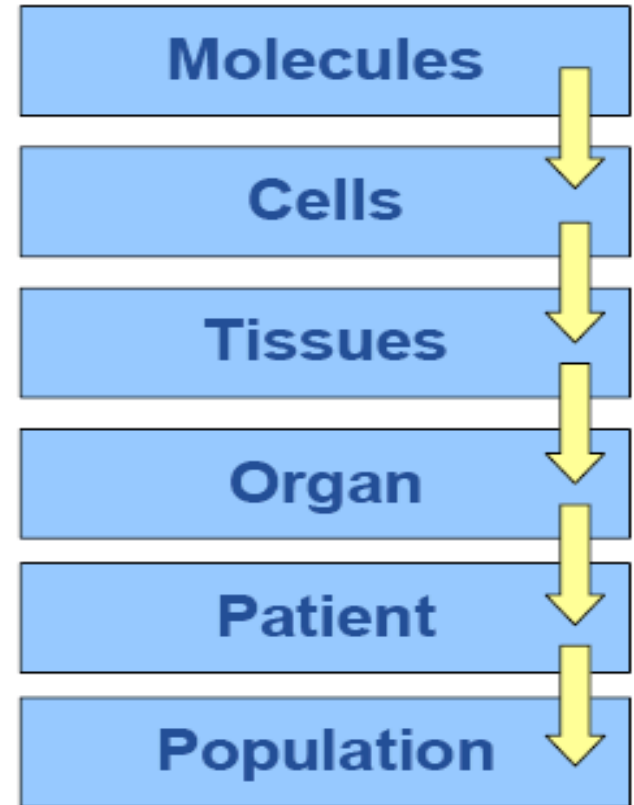
→ **Selectivity: 1 in 10^{13}**

(Like looking for a needle in 20 million haystacks)

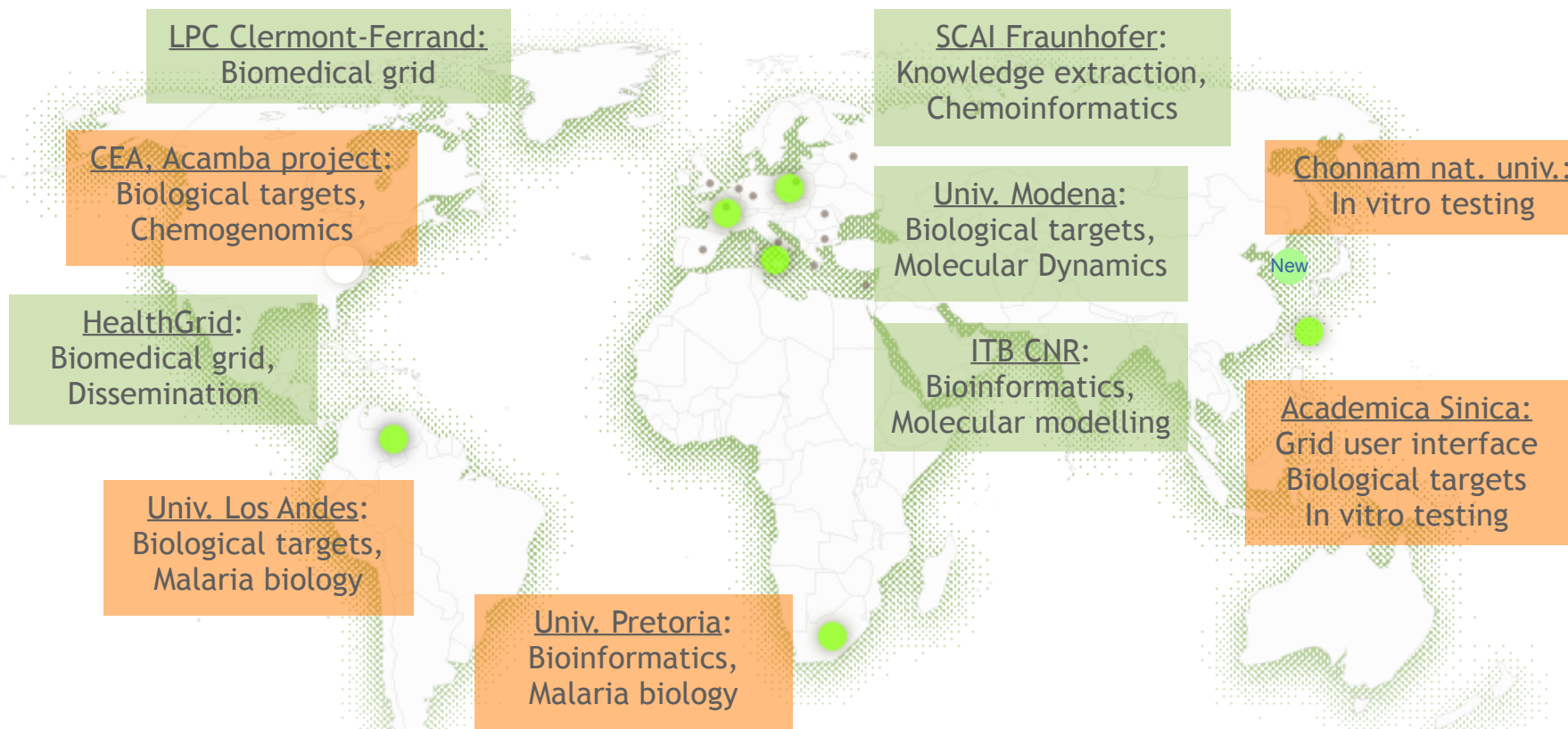
- **Bioinformatics**
 - Genomics
 - Proteomics
 - Phylogeny...

- **Medical imaging**
 - Medical imaging
 - Computer Aided Diagnosis
 - Therapy planning
 - Simulation...

- **Life sciences**
 - Drug discovery
 - Epidemiology
 - ...

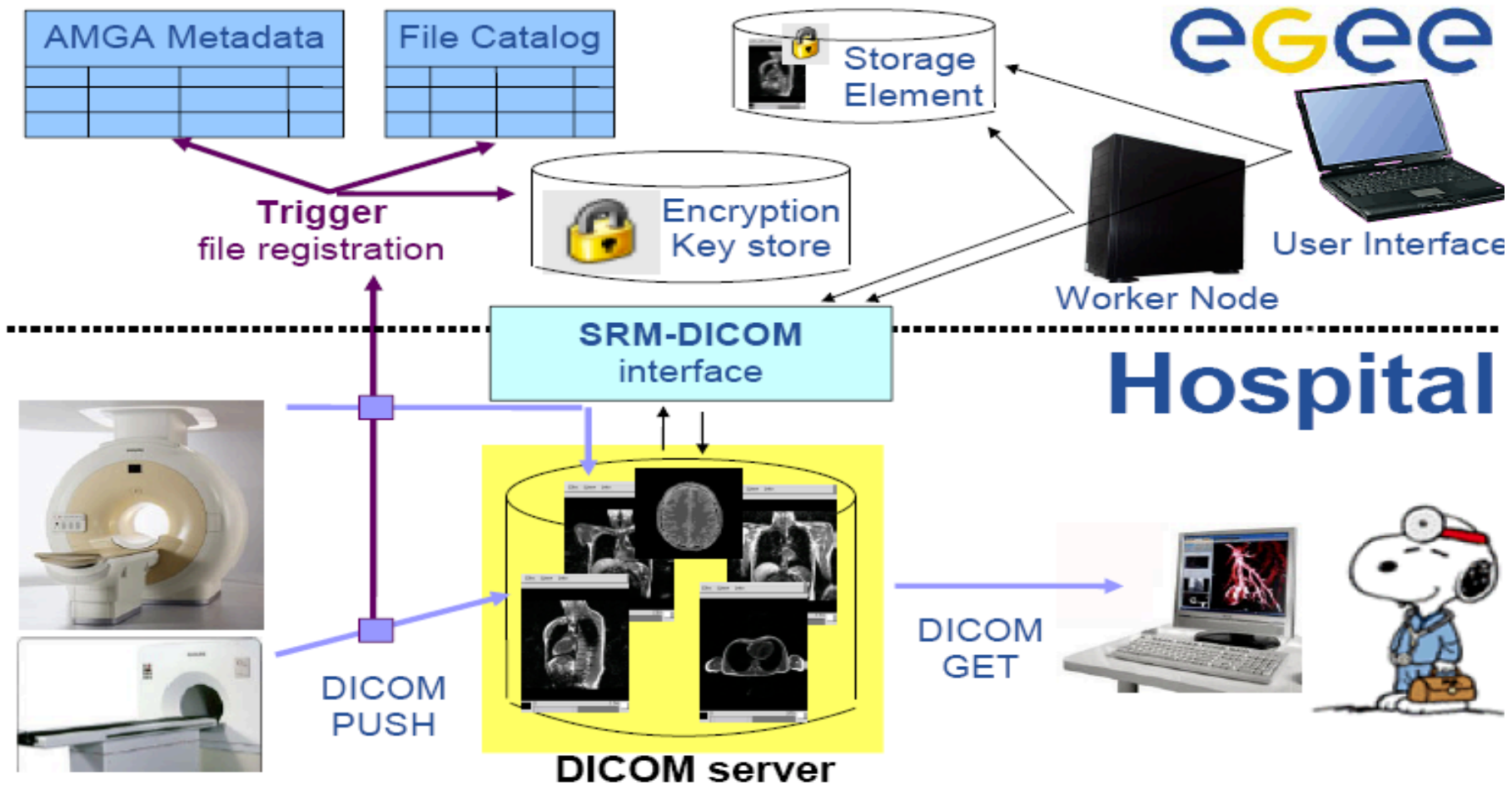


Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin



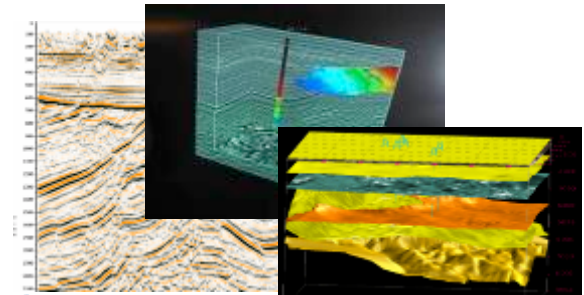
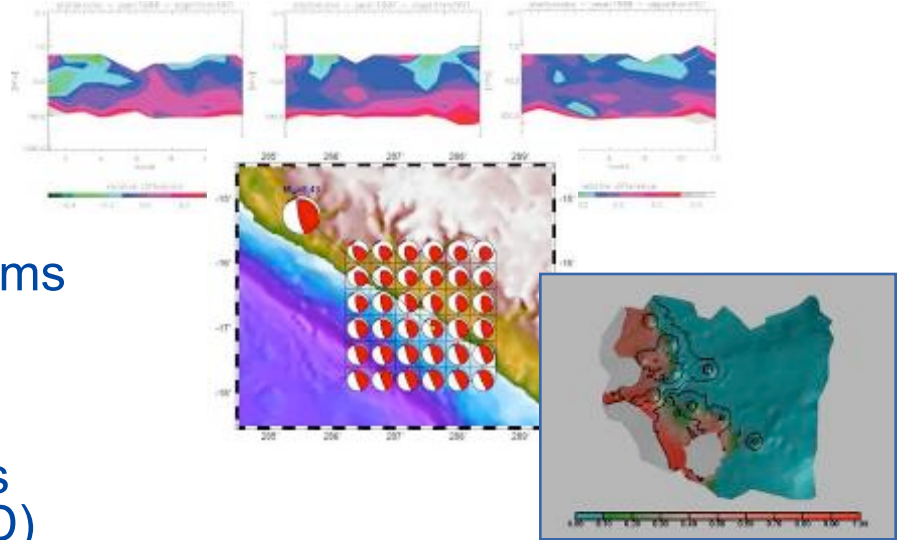
Avian flu data challenge: in the selection of 2250 compounds out of initial 308585 compounds, an enrichment factor of 111 was observed. Experimental trial confirms 7 actives out of 123 tested gave “potential hits”.

Data challenges on malaria: the 25 most promising compounds out of 500.000 are now being tested in vitro at Chonnam National University



Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin

- **Earth Observations by Satellite**
 - Ozone profiles
 - **Solid Earth Physics**
 - Fast Determination of mechanisms of important earthquakes
 - **Hydrology**
 - Management of water resources in Mediterranean area (SWIMED)
 - **Geology**
 - Geocluster: R&D initiative of the Compagnie Générale de Géophysique
- A large variety of applications ported on EGEE



Laboratory
Measurements
DB

GeoArchaeology
DB

Archaeo
Climatology
DB

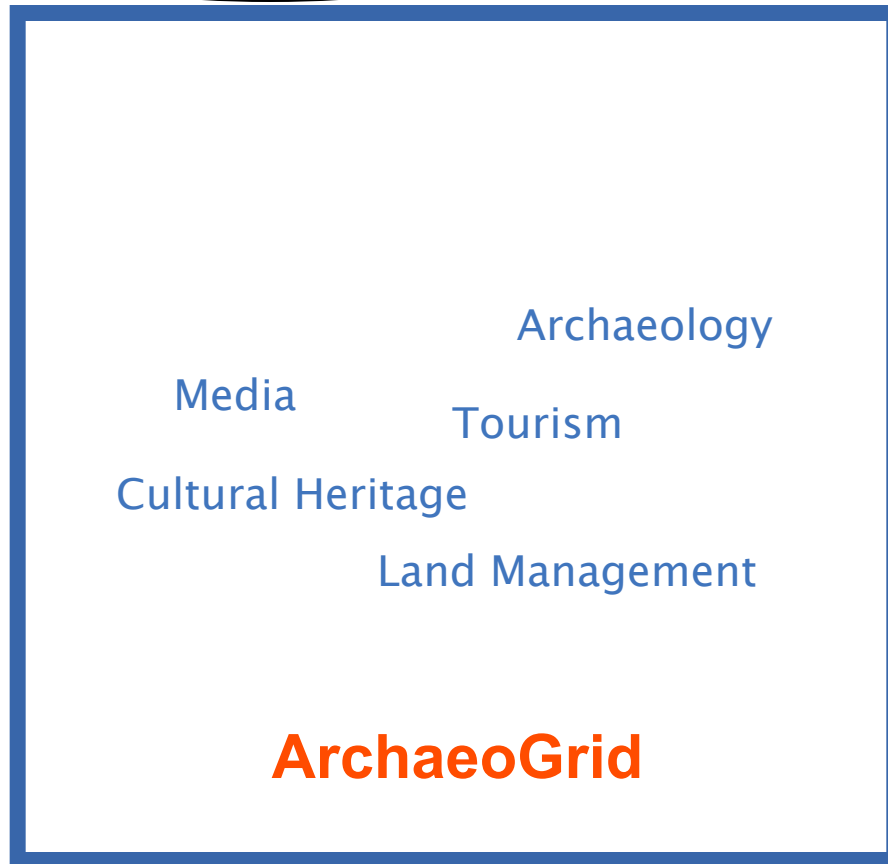
Archaeo
Zoology/Botanic
DB

Archaeological
bibliography
DB

Archaeological
Objects
DB

Images
DB

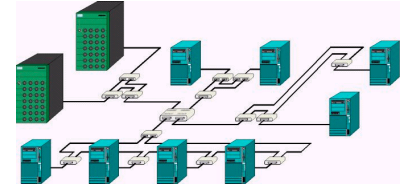
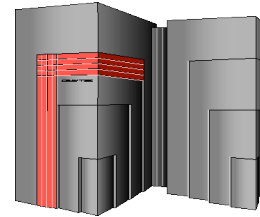
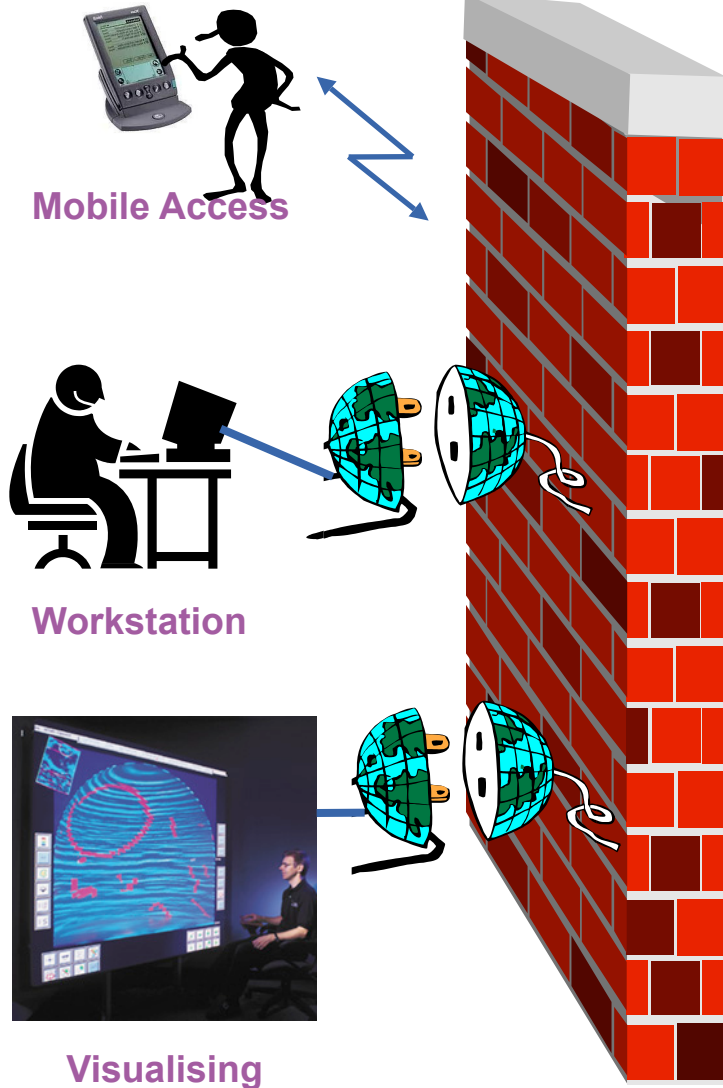
TextFile
DB



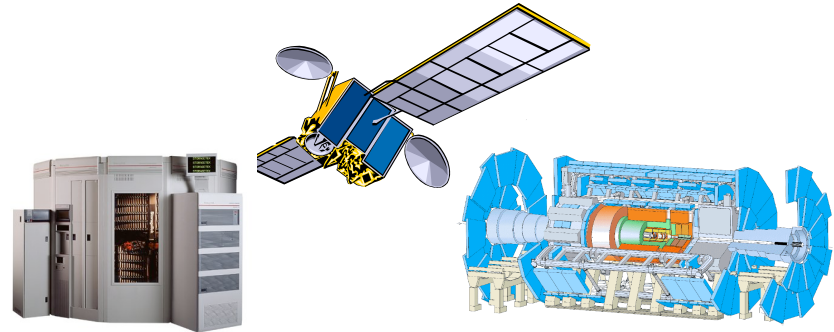
Simulation/VR
DB

Archaeological
GIS

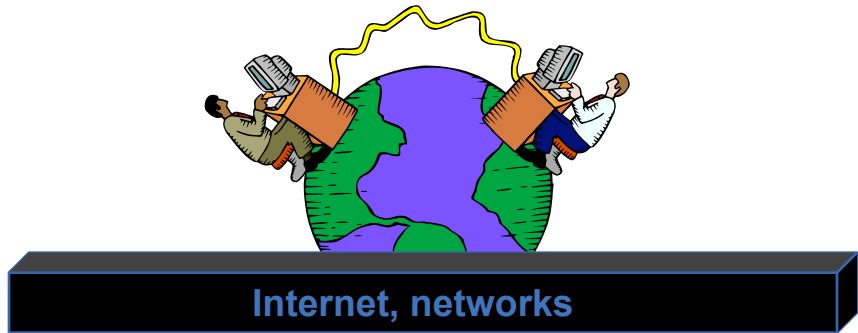
Grid concepts



Supercomputer, PC-Cluster



Data-storage, Sensors, Experiments

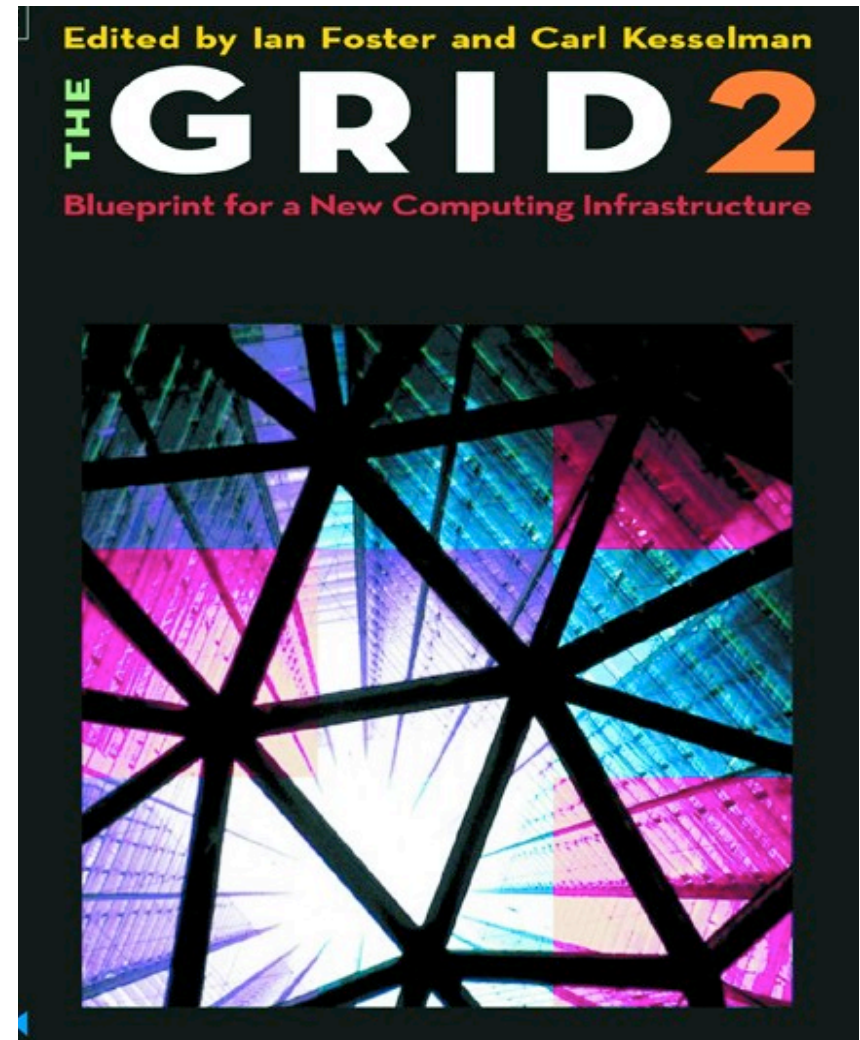


Internet, networks

GRID MIDDLEWARE

- The grid vision is of “Virtual computing” (+ information services to locate computation, storage resources)
 - Compare: The web: “virtual documents” (+ search engine to locate them)

- **MOTIVATION: collaboration through sharing resources (and expertise) to expand horizons of**
 - Research
 - Commerce – engineering, ...
 - Public service – health, environment,...



- Enabling a whole-system approach
- A challenge to the imagination
- Effect > Σparts

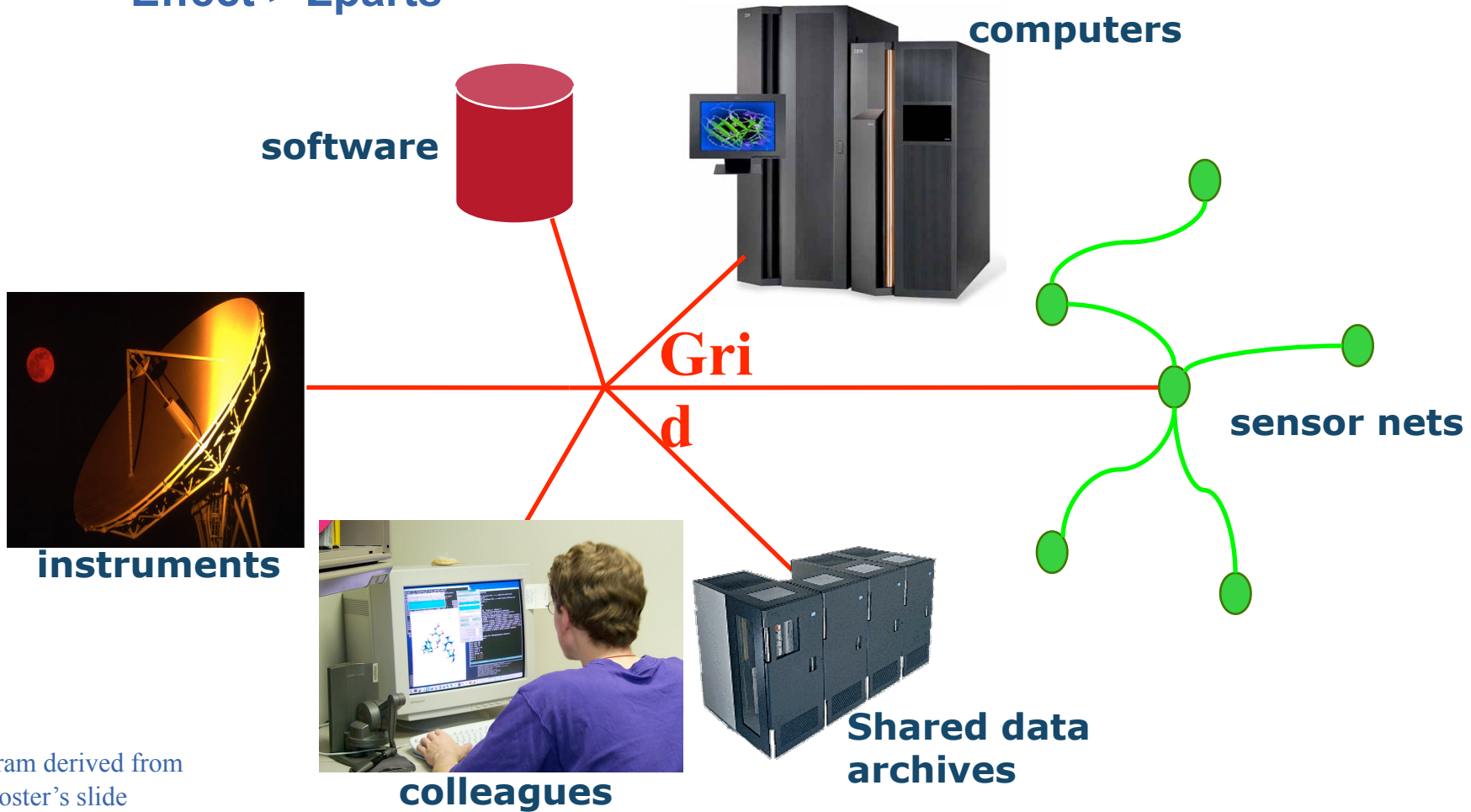
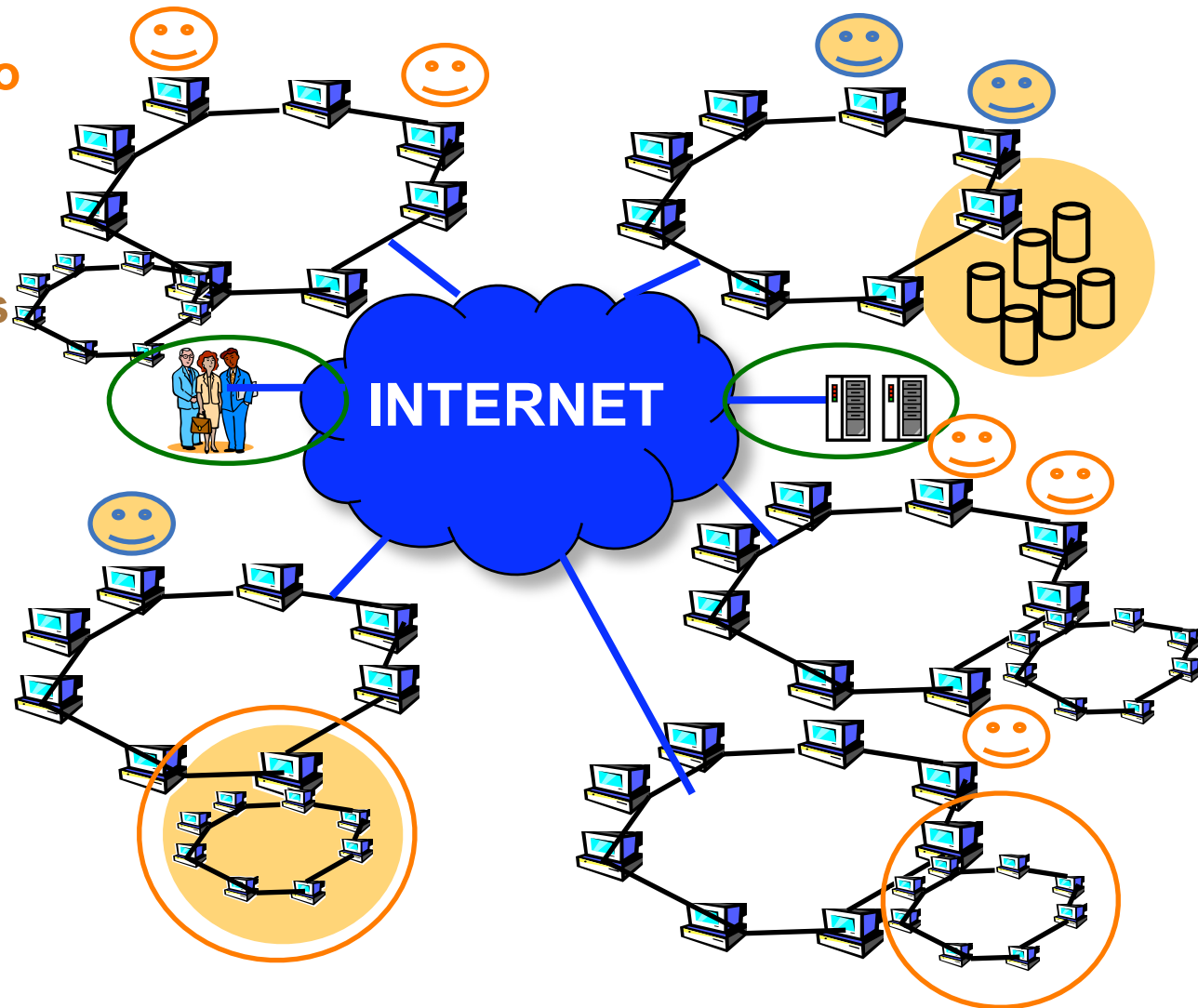


Diagram derived from Ian Foster's slide

- **Flexible, simplified orchestration of resources available to a collaboration**
 - Across administrative domains
 - Abstractions hide detail of individual resources
 - Conform to Grid’s procedures to gain benefit
 - Operations services (people and software)
- **Increased utilisation**
 - A collaboration shares its resources building on Grid services
 - Collaborations share resources
 - Each contributes average requirements (cpus, storage)
 - Each can benefit from
 - *Heterogeneity*
 - *Scale*

- **What is a Virtual Organisation?**
 - People in different organisations seeking to cooperate and share resources across their organisational boundaries
 - E.g. A research collaboration
- **Each grid is an infrastructure enabling one or more “virtual organisations” to share and access resources**
- **Each resource is exposed to the grid through an abstraction that masks heterogeneity, e.g.**
 - Multiple diverse computational platforms
 - Multiple data resources
- **Resources are usually owned by VO members. Negotiations lead to VOs sharing resources**

- **Virtual organisations negotiate with sites to agree access to resources**
- **Grid middleware runs on each shared resource to provide**
 - Data services
 - Computation services
 - Single sign-on
- **Distributed services (both people and middleware) enable the grid**



- **When using a Grid you**

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**

- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**

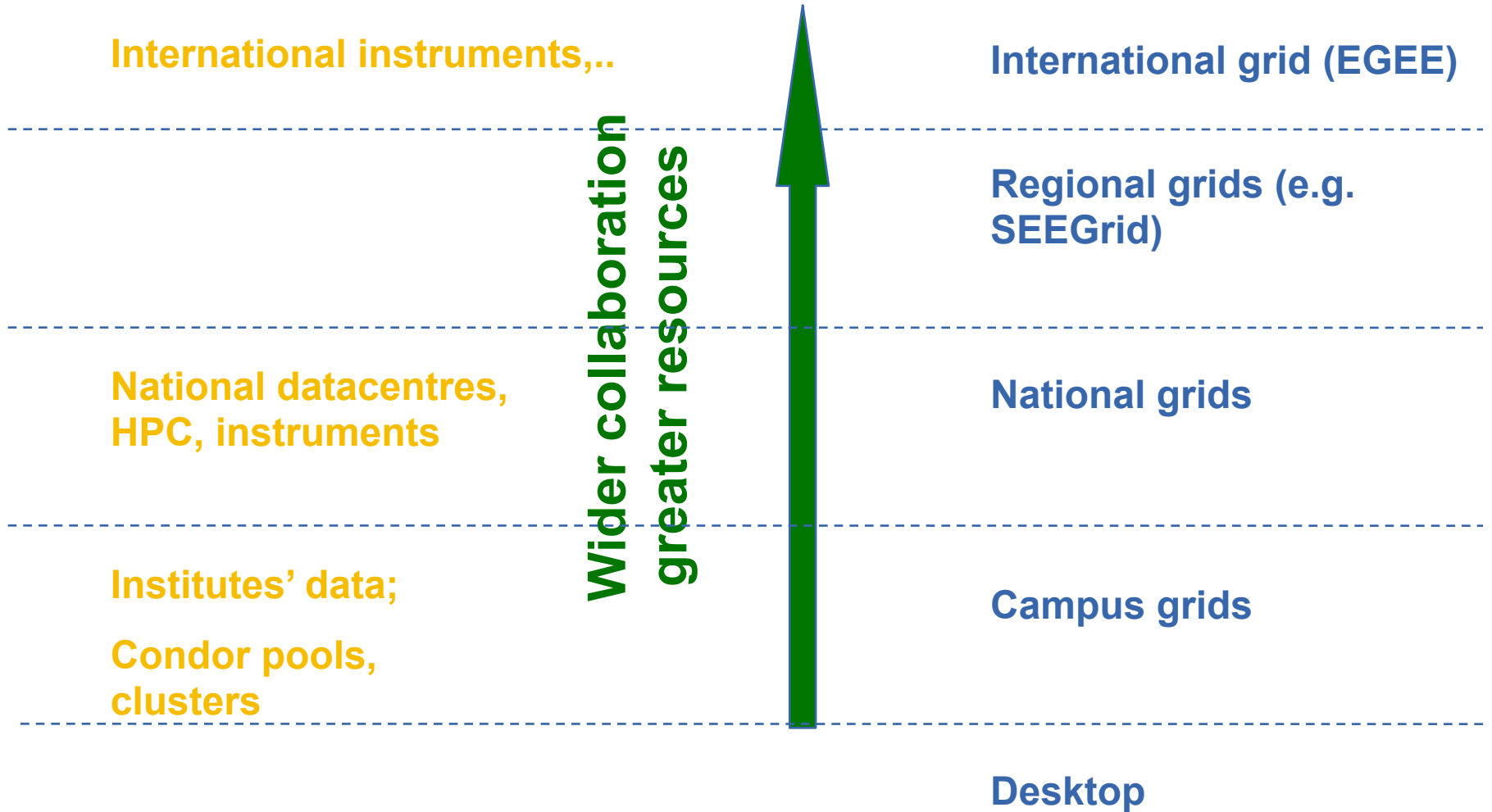
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

- **When using a PC or workstation you**
 - Login with a username and password (“Authentication”)
 - Use rights given to you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

- **When using a PC or workstation you**
 - Login with a username and password (“Authentication”)
 - Use rights given to you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

- **When using a PC or workstation you**
 - Login with a username and password (“Authentication”)
 - Use rights given to you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **Operating system**
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

- **When using a PC or workstation you**
 - Login with a username and password (“Authentication”)
 - Use rights given to you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **Operating system**
- **One admin. domain**
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

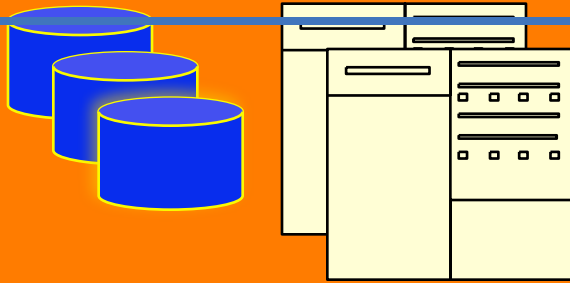


Application

Application
toolkits

Higher-level grid
services (brokering,...)

Basic Grid services:
AA, job submission, info, ...

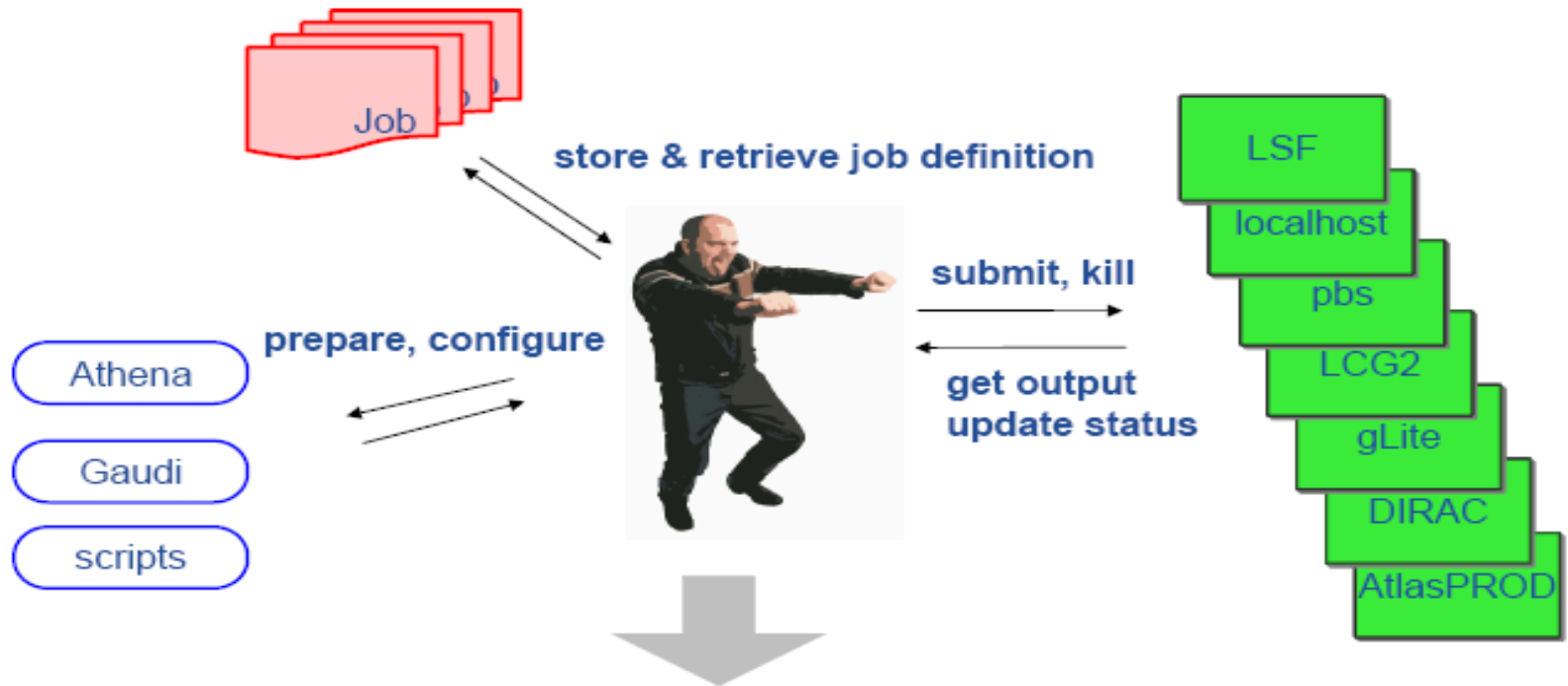


Where computer science meets the
application communities!

High level tools and
VO-specific developments:

- Portals
- Virtual Research Environments
- Semantics, ontologies
- Workflow
- Registries of VO services

Production grids provide these
services.



Ganga is:

Ganga4

+ split, merge, monitor

- a lightweight user tool
 - a developer framework
- <http://ganga.web.cern.ch/>

If "The Grid"
vision leads us
here...





If "The Grid"
vision leads us
here...

... then where are
we now?



If "The Grid"
vision leads us
here...

... then where are
we now?

Where are we now? –user's view



Research

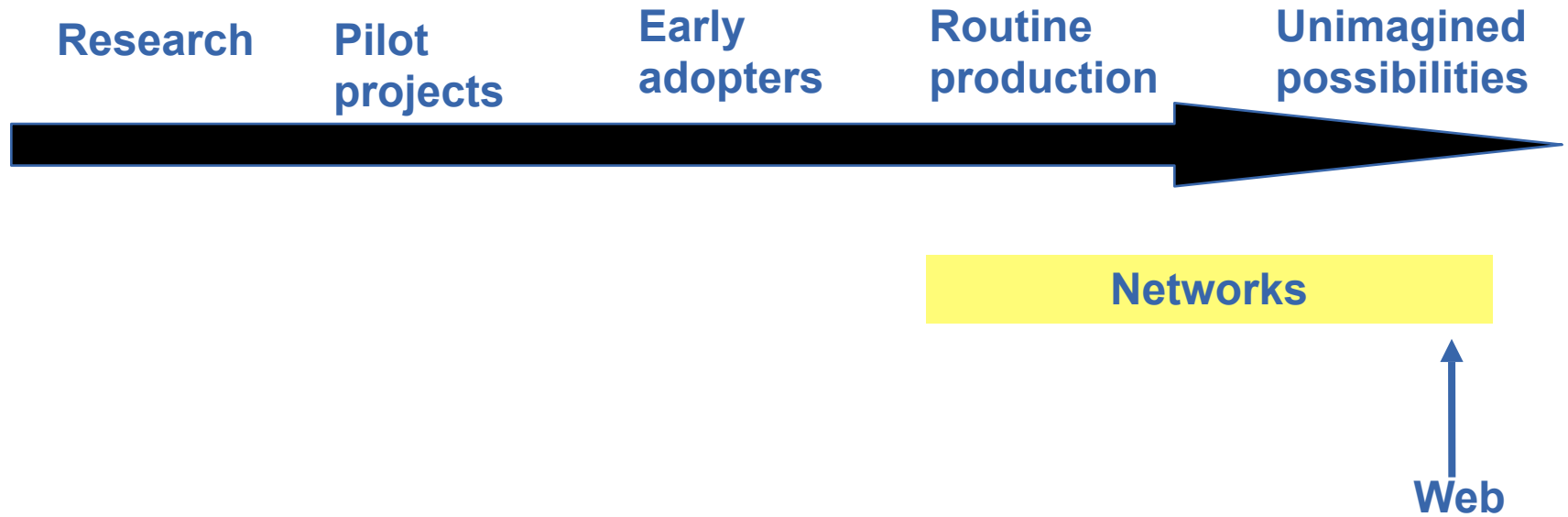
Pilot
projects

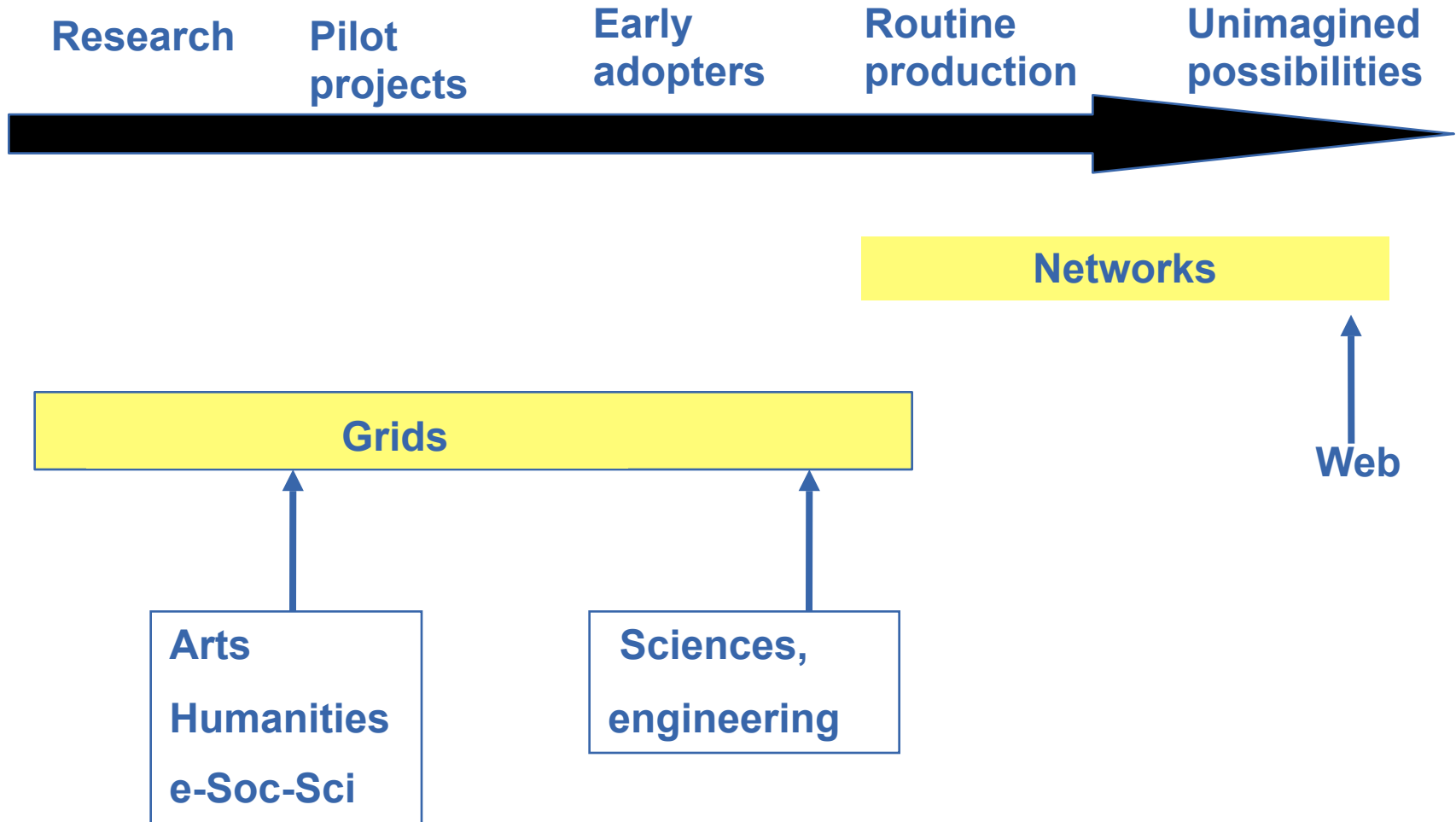
Early
adopters

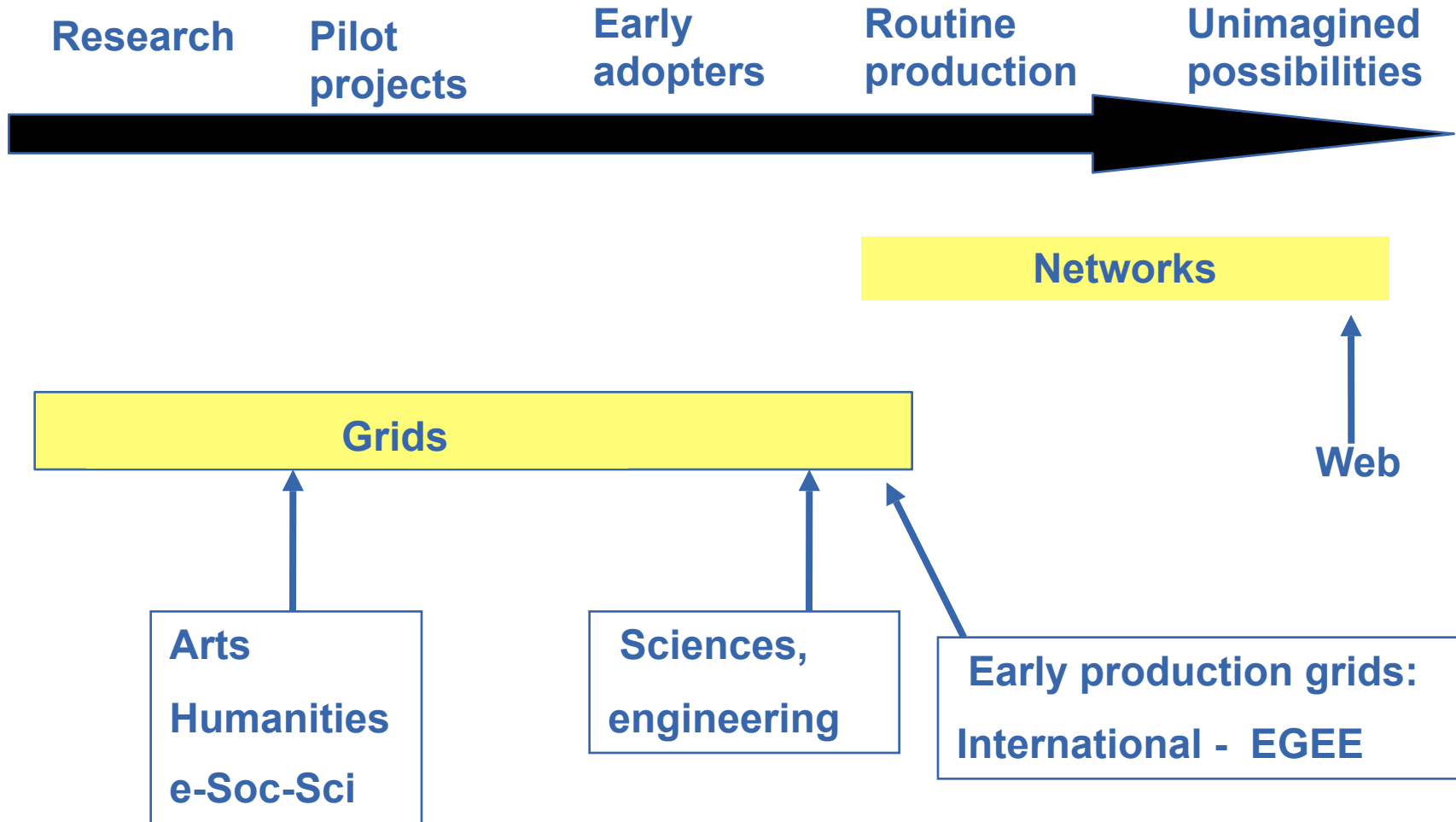
Routine
production

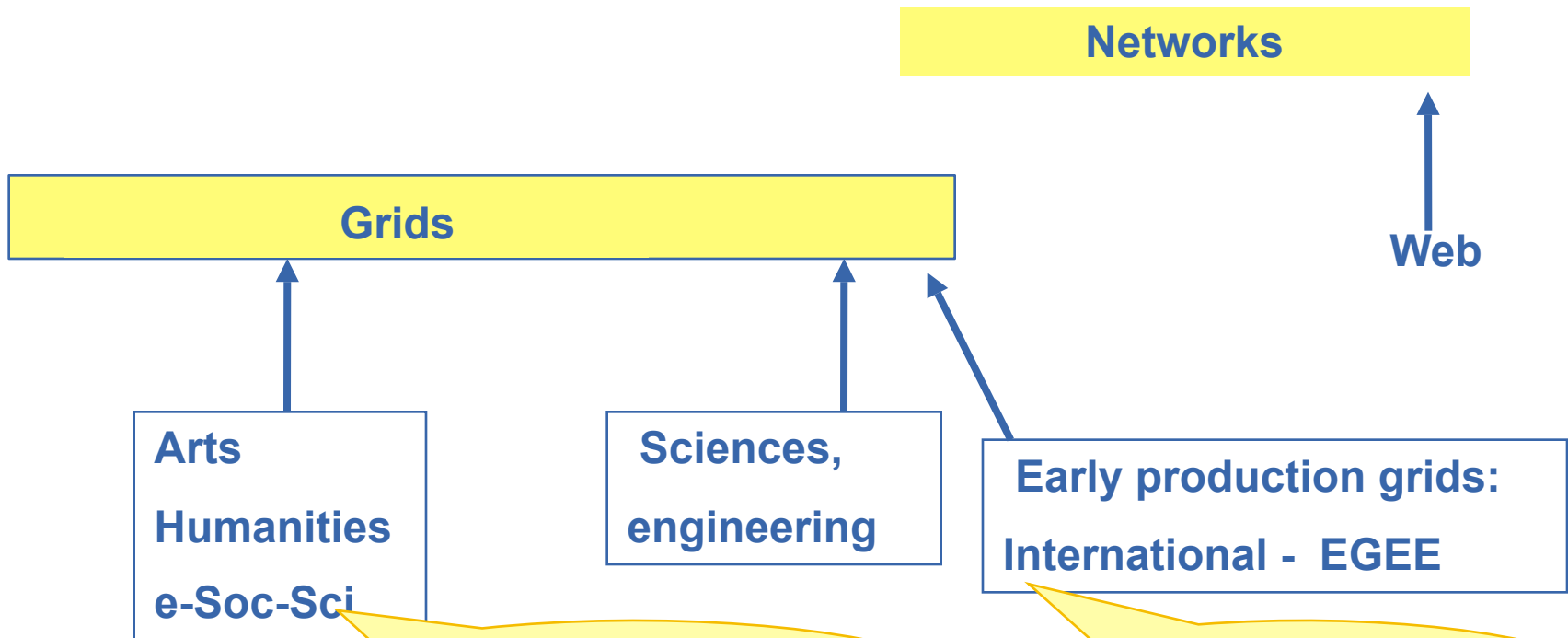
Unimagined
possibilities

Where are we now? –user's view









Service-oriented, workflow, “legacy” data

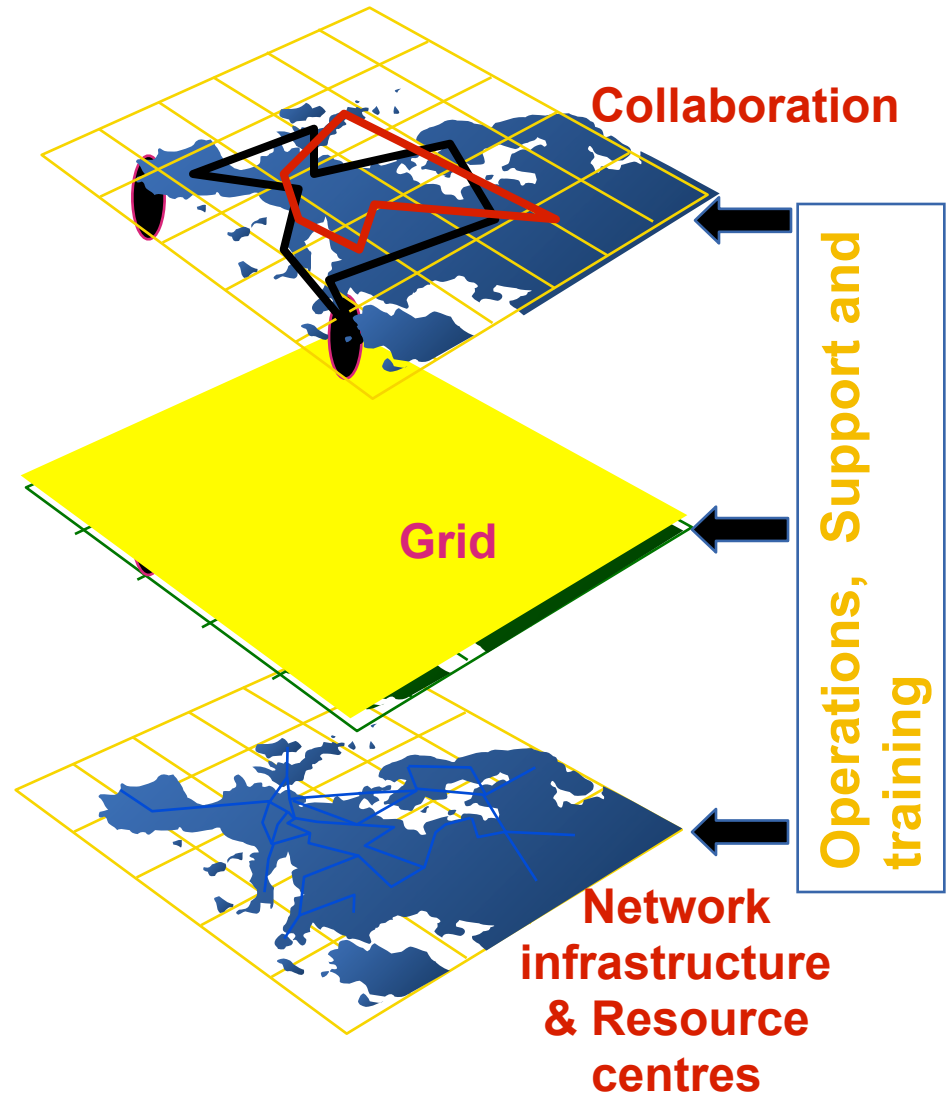
High throughput, new data

Types of use:

- Many key concepts identified and known
- Many grid projects have tested, and benefit from, these
 - Empowering collaborations
 - Resource-sharing
- Major efforts now on establishing:
 - **Production Grids *for multiple VO's***
 - “Production” = Reliable, sustainable, with commitments to quality of service
 - Each has
 - *One stack of middleware that serves many research communities*
 - *Establishing operational procedures and organisation*
 - Challenge for EGEE-II: federate these!
 - **Standards** (a slow process)
 - e.g. Open (formerly Global) Grid Forum, <http://www.gridforum.org/>
 - Extending web services
 - **Broadening range of research communities**
 - arts and humanities, social science ...

- **Providers of resources (computers, databases,..) need risks to be controlled: they are asked to trust users they do not know**
- **User's need**
 - single sign-on: to be able to logon to a machine that can pass the user's identity to other resources
 - To trust owners of the resources they are using
- **Build middleware on layer providing:**
 - *Authentication*: know who wants to use resource
 - *Authorisation*: know what the user is allowed to do
 - *Security*: reduce vulnerability, e.g. from outside the firewall
 - *Non-repudiation*: knowing who did what
- **The “Grid Security Infrastructure” middleware is the basis of (most) production grids**

- **Grids enable virtual computing across administrative domains**
 - Resources share authorisation and authentication
 - Resources accessed thru abstractions
- **Motivations:**
 - Collaborative research, diagnostics, engineering, public service,...
 - Resource utilisation and sharing



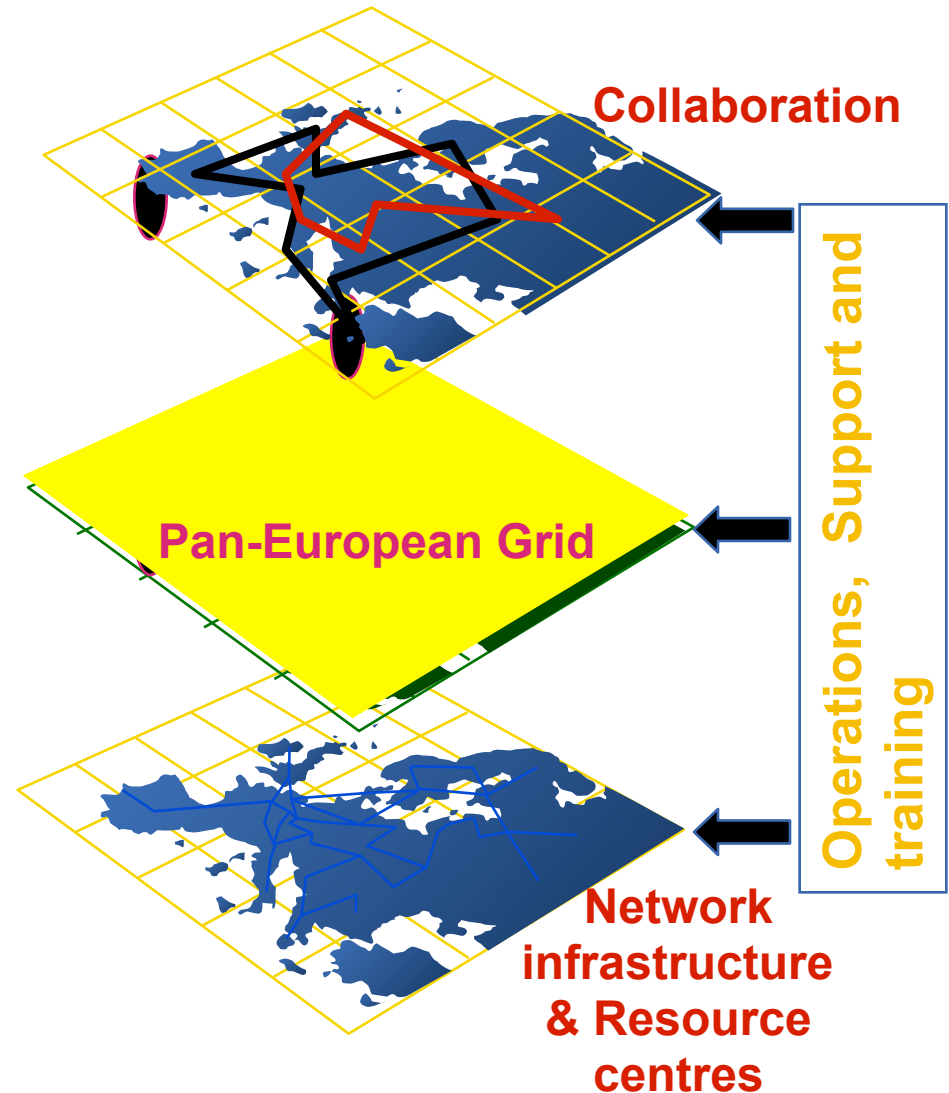
- **Open Grid Forum** <http://www.ggf.org/>
- **The Grid Cafe** www.gridcafe.org
- **Grid Today** <http://www.gridtoday.com/>
- **Globus Alliance** <http://www.globus.org/>
- **International Science Grid This Week** <http://www.isgtw.org/>
- **International School of Grid Computing** <http://www.issgc.org/> near Stockholm, from 8th to 20th July 2007.

- **What is EGEE?**
 - Goals
 - Status
 - Activities
- **Grid services: gLite 3.0**
- **Sources of further information**



A four year programme:

- Build, deploy and operate a consistent, robust a large scale production grid service that
 - Links with and build on national, regional and international initiatives
- Improve and maintain the middleware in order to deliver a reliable service to users
- Attract new users from research and industry and ensure training and support for them

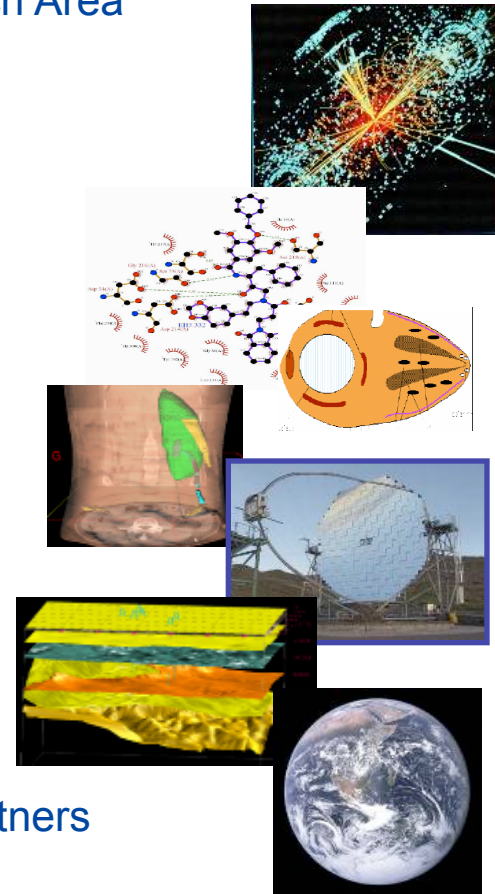


- **Infrastructure**
 - Manage and operate production Grid for European Research Area
 - Interoperate with e-Infrastructure projects around the globe
 - Contribute to Grid standardisation efforts

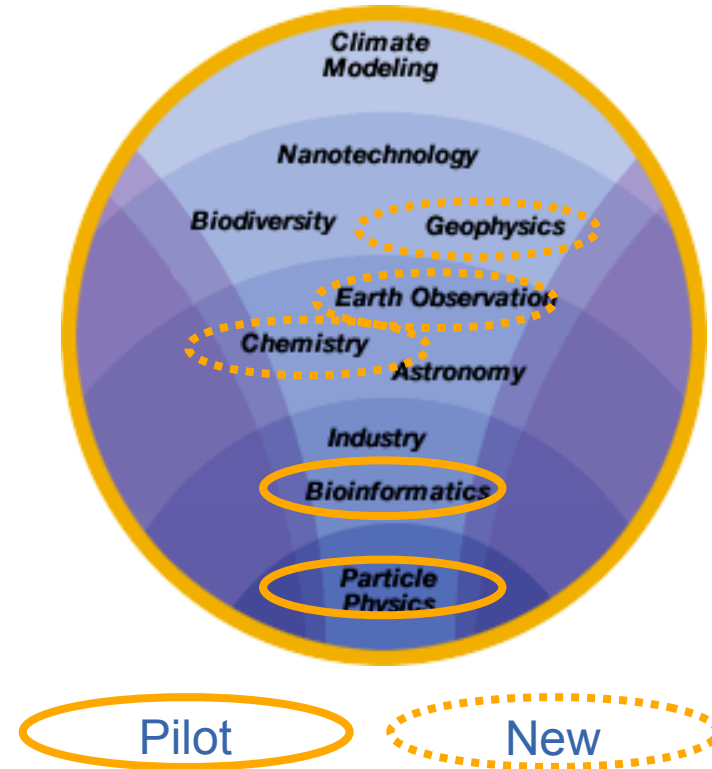
- **Support applications from diverse communities**
 - High Energy Physics
 - Biomedicine
 - Earth Sciences
 - Astrophysics
 - Computational Chemistry
 - Fusion
 - Geophysics
 - Finance, Multimedia
 - ...

- **Business**
 - Forge links with the full spectrum of interested business partners

- + Disseminate knowledge about the Grid through training
- + Prepare for sustainable European Grid Infrastructure



- **Established production quality sustained Grid services**
 - 3000 users from at least 5 disciplines
 - Goal was to integrate 50 sites into a common infrastructure → currently 180
 - offer 5 Petabytes (10^{15}) storage
- **Demonstrated a viable general process to bring other application communities on board**
- **Secured a second phase from April 2006**





- Flagship grid infrastructure project co-funded by the European Commission
- Now in 2nd phase with 91 partners in 32 countries

Objectives

- Large-scale, production-quality grid infrastructure for e-Science
- Attracting new resources and users from industry as well as science
- Maintain and further improve gLite Grid middleware



EGEE
Enabling Grids
for E-science

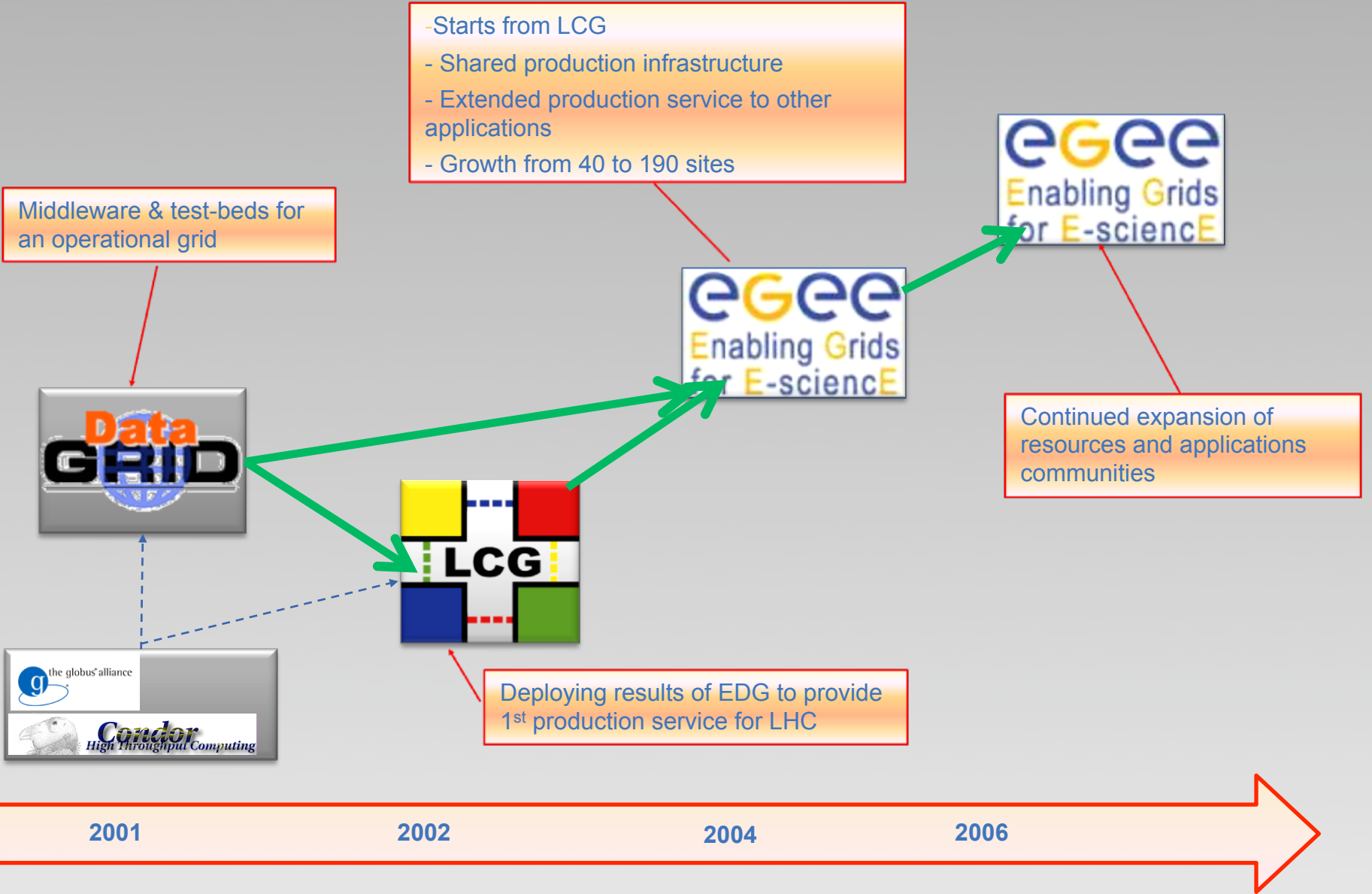


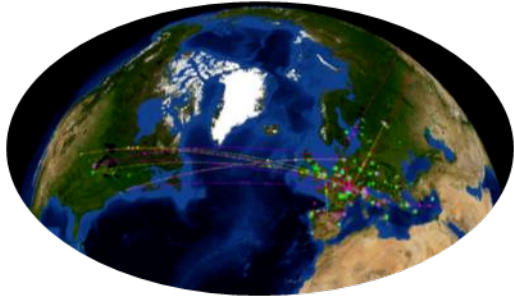
Scheduled = 15725
Running = 8887

13:24:23 UTC



Enabling Grids for E-science





Test-beds & Services

Production Service

Pre-production service

Certification test-beds

Training infrastructure

Support Structures & Processes

Operations Coordination Centre

Regional Operations Centres

Global Grid User Support

EGEE Network Operations Centre

Operational Security Coordination Team

Training activities

Security & Policy Groups

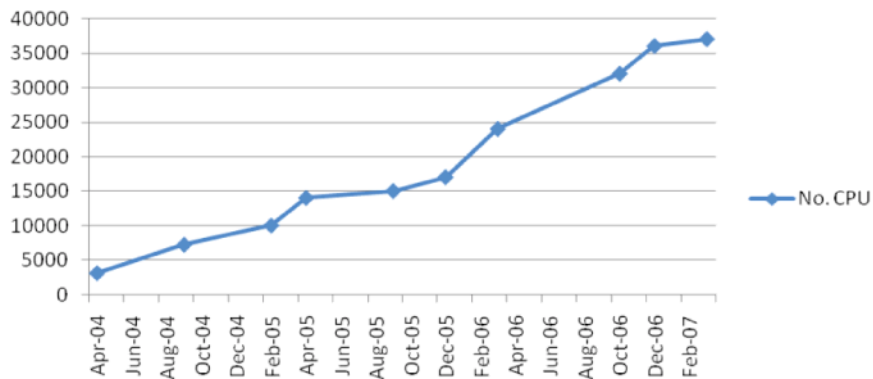
Joint Security Policy Group

EuGridPMA (& IGTF)

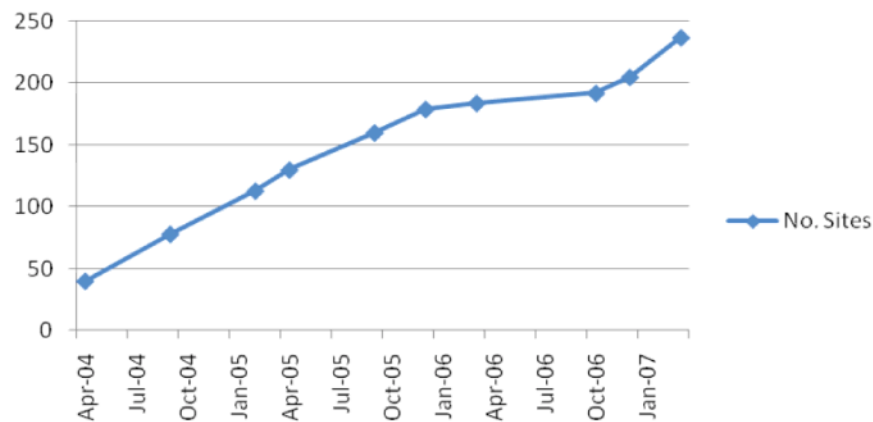
Grid Security Vulnerability Group

Operations Advisory Group

No. CPU



No. Sites

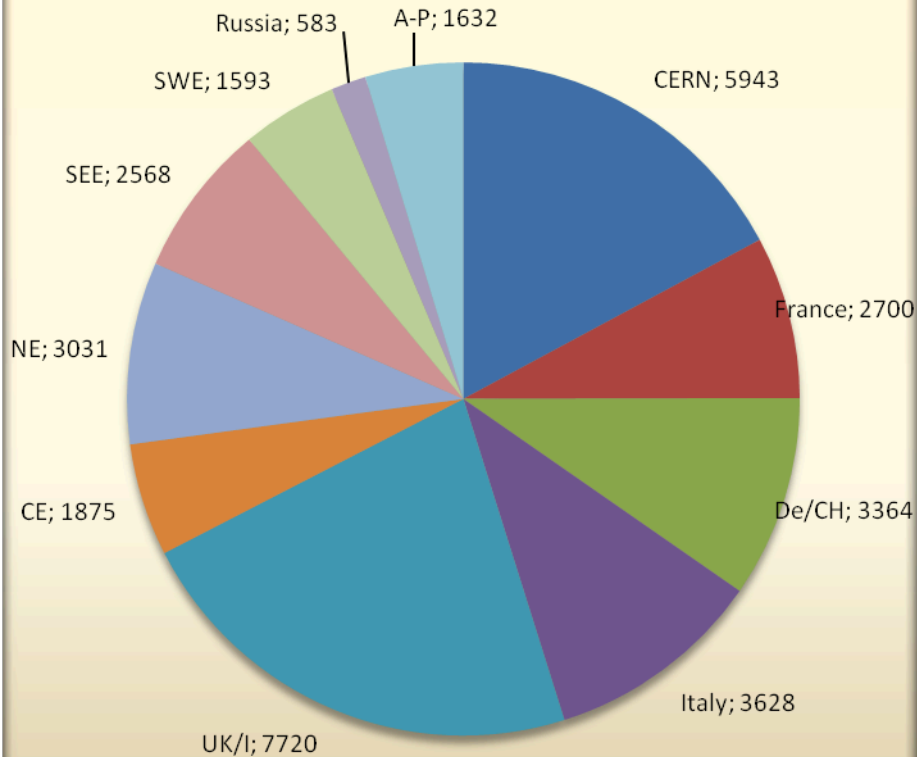


Scheduled = 15725
Running = 8887

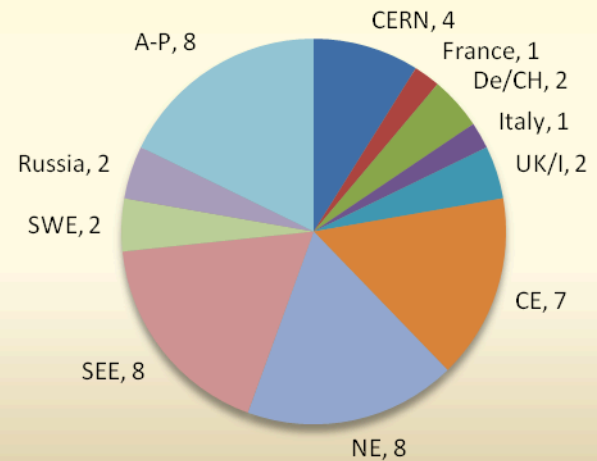


GridPP
UK Computing for Particle Physics

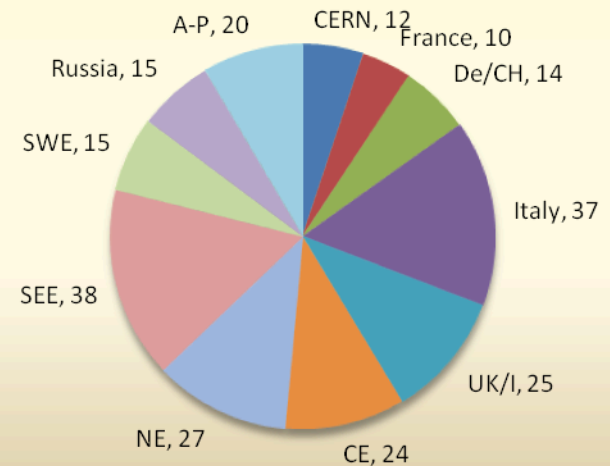
CPU / ROC



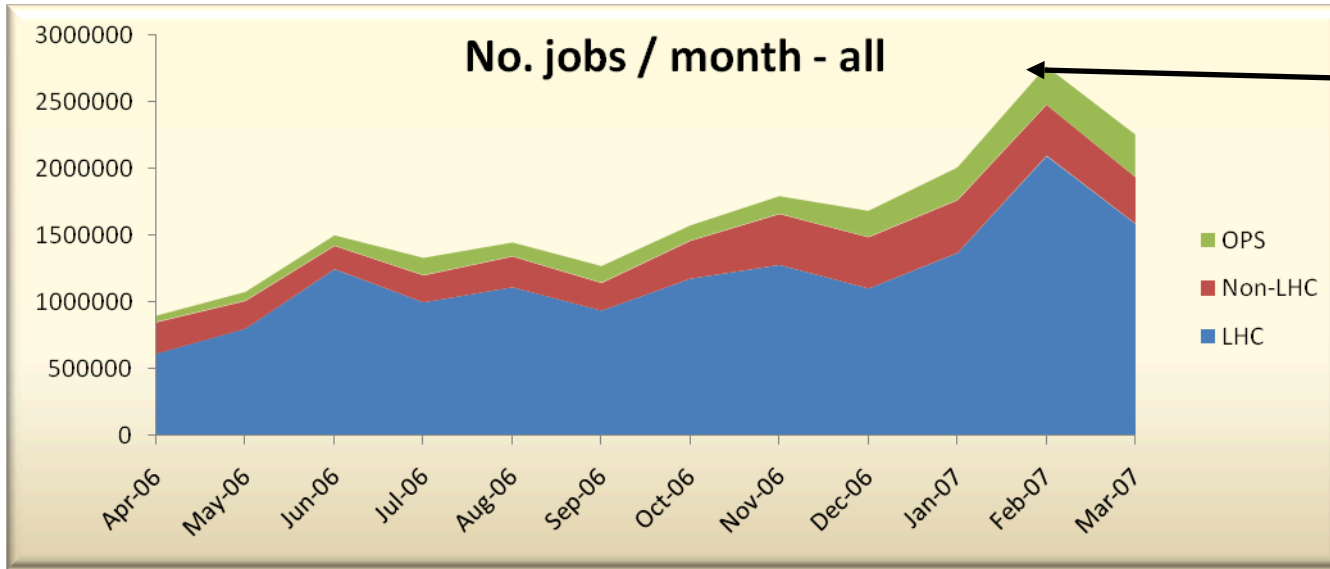
Countries / ROC



Sites / ROC

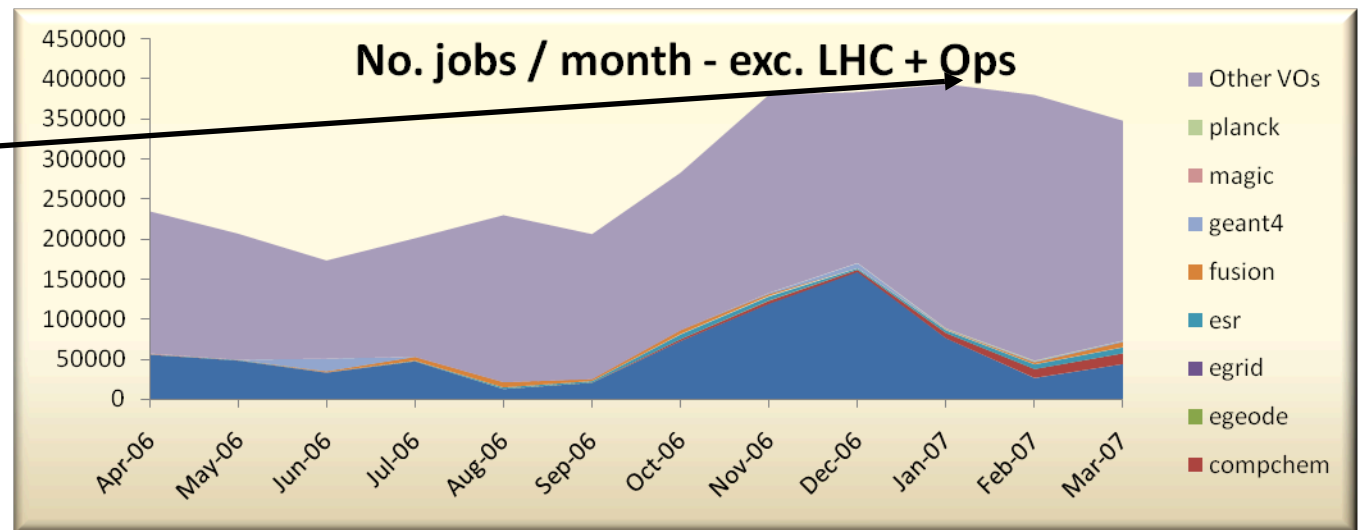


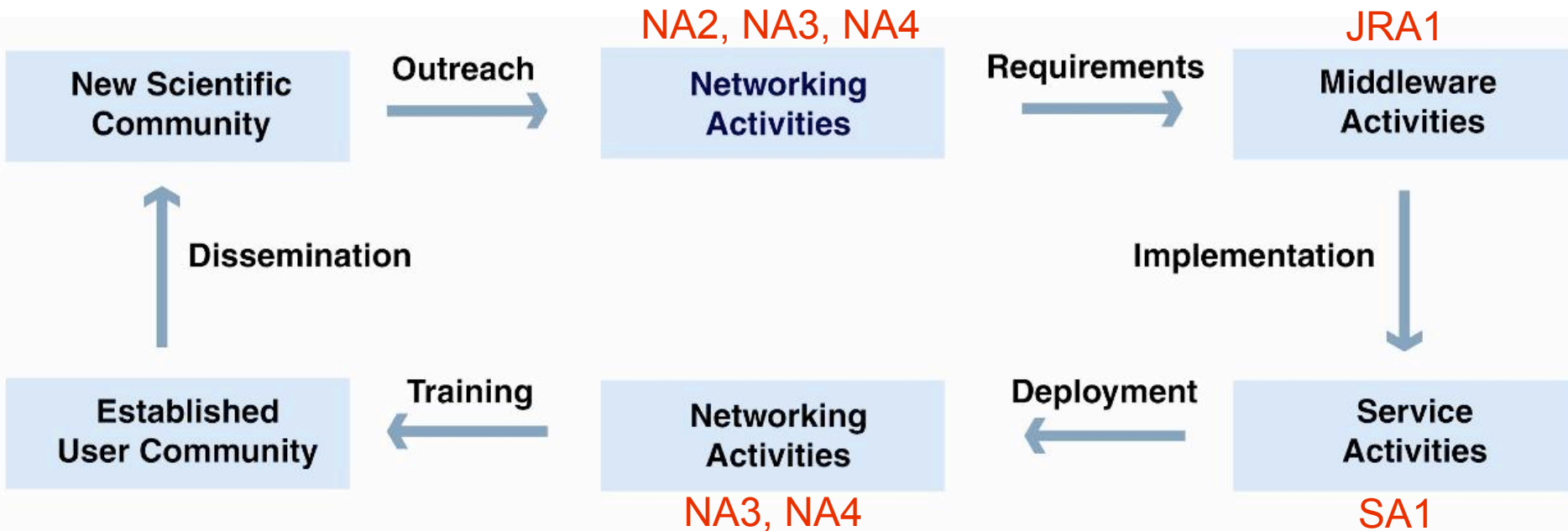
- ▶ 35000 CPU
- ▶ 45 countries (31 partner countries)
- ▶ 237 sites (131 partner sites)



98000 jobs/day

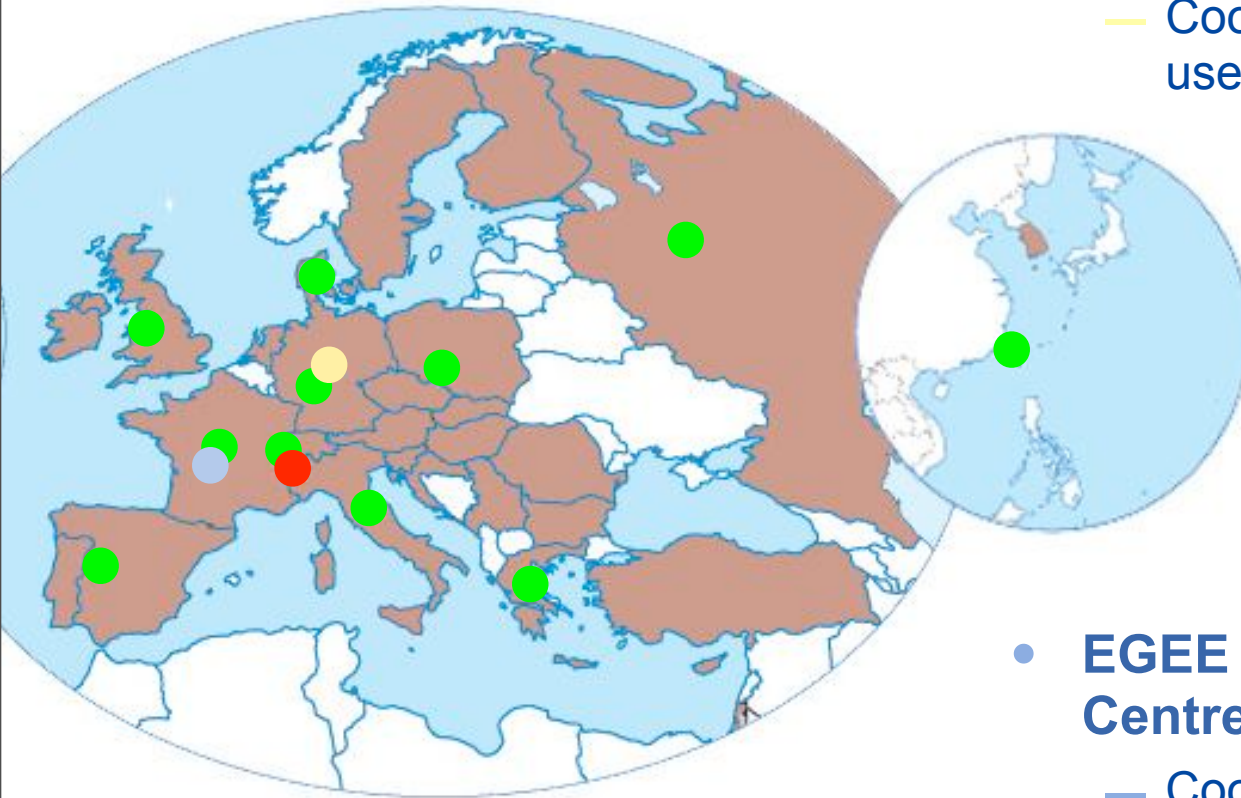
13000 jobs/day



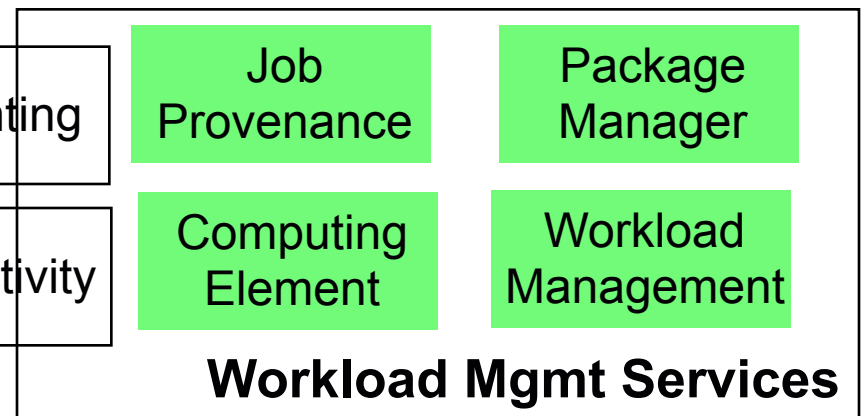
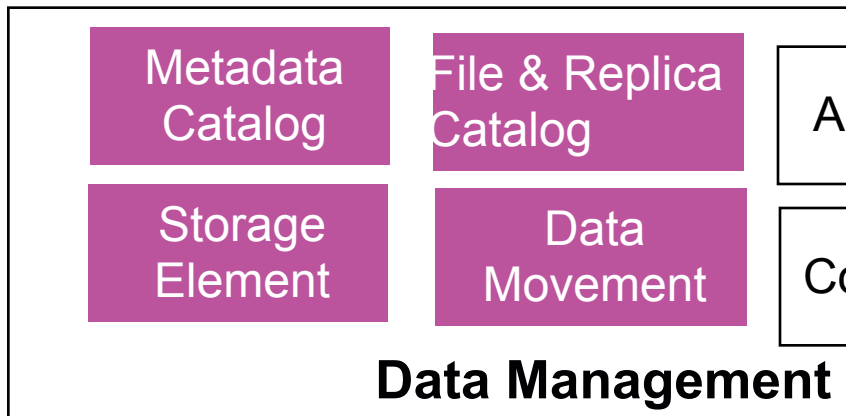
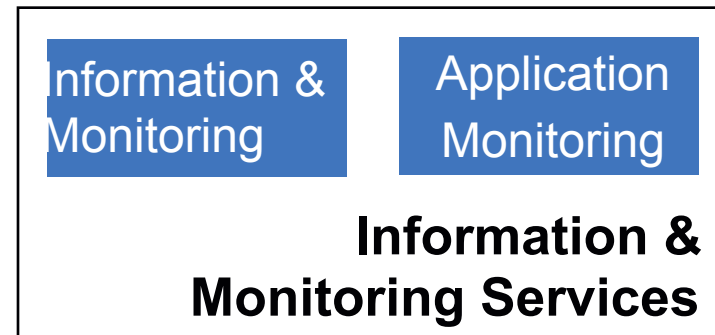
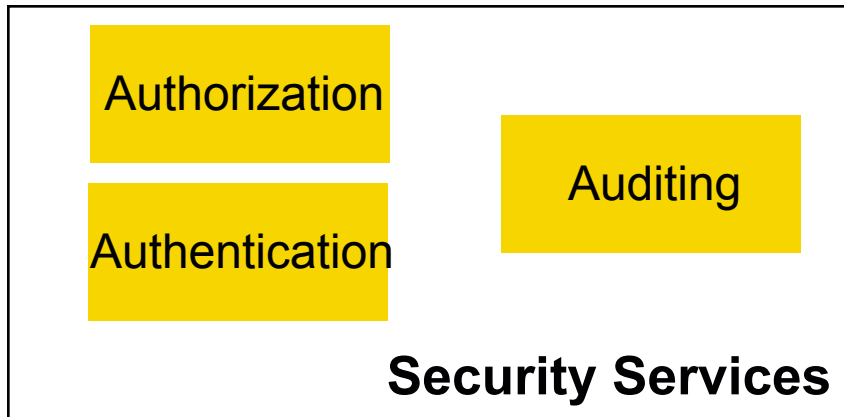
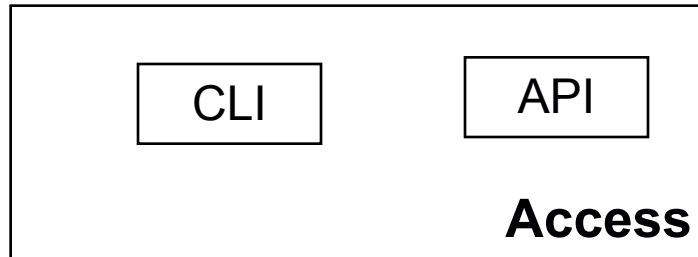


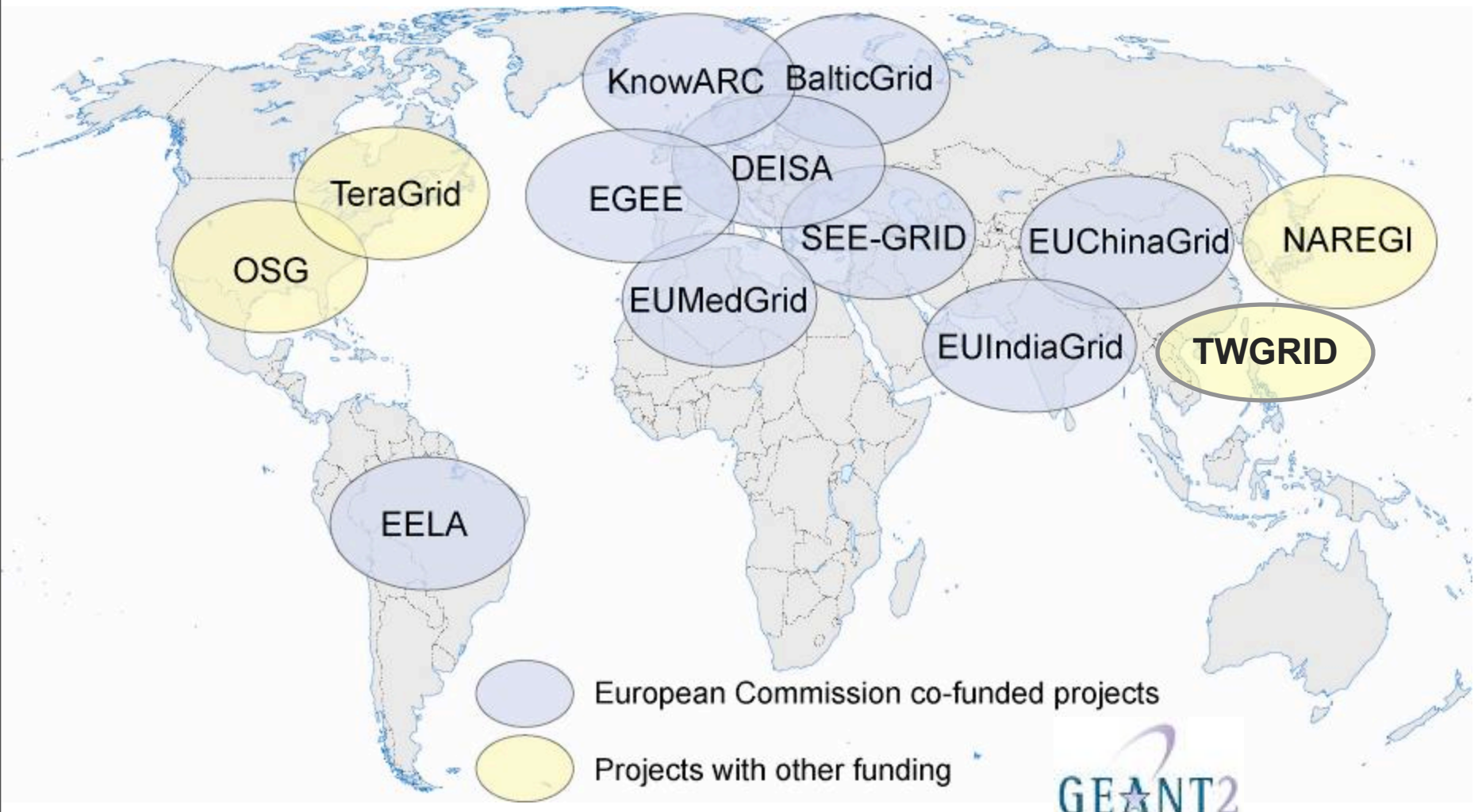
Building effective user communities

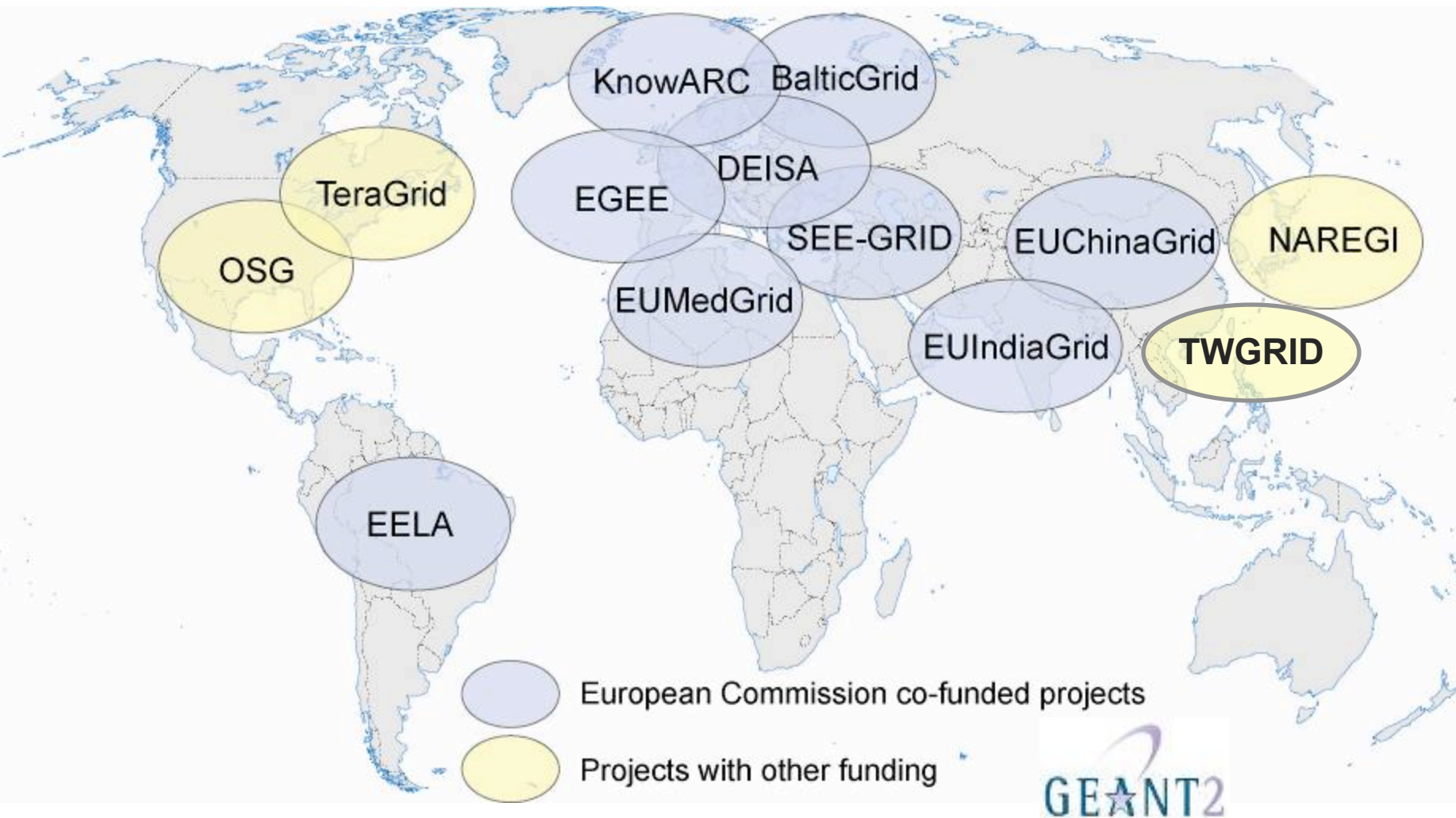
- **Operations Coordination Centre**
 - Management, oversight, coordination
- **Regional operations Centres**
 - Core support infrastructure
- **Grid User Support (GGUS)**
 - Coordination, management of user support



- **EGEE Network Operations Centre (ENOC)**
 - Coordination with NRENs & GEANT2

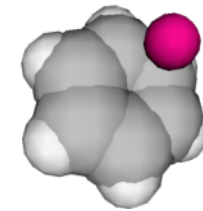
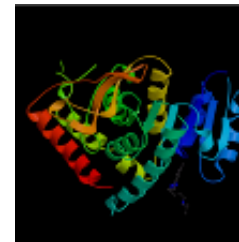




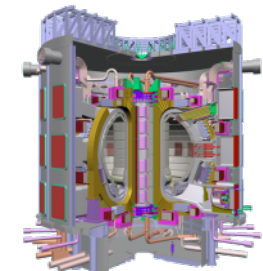
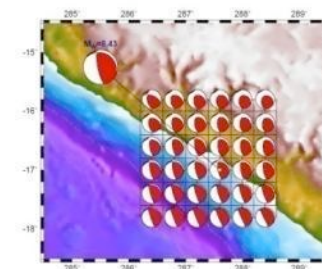


Potential for linking ~80 countries

- Disciplines: 10
- Sub-disciplines: 36
- See growth and diversification of applications.
- Reported apps. only



| | PM3 | PM11 |
|--------------------------|-----------|------------|
| Astronomy & Astrophysics | 2 | 8 |
| Computational Chemistry | 6 | 27 |
| Earth Science | 16 | 16 |
| Fusion | 2 | 3 |
| High-Energy Physics | 9 | 11 |
| Life Sciences | 23 | 39 |
| Others | 4 | 14 |
| Total | 62 | 118 |



Condensed Matter Physics
 Comp. Fluid Dynamics
 Computer Science/Tools
 Civil Protection

- **EGEE Conference: 25-29 September 2006**
<http://www.eu-egee.org/news/registration-open-for-egee201906-conference-September-2006-geneva/>
- **EGEE digital library:** <http://egee.lib.ed.ac.uk/>
- **EGEE** www.eu-egee.org
- **EGEE: 1st user Forum**
<http://egee-intranet.web.cern.ch/egee-intranet/User-Forum>
- **gLite** <http://www.glite.org/>
- **Open Grid Forum** <http://www.gridforum.org/>
- **Globus Alliance** <http://www.globus.org/>
- **VDT** <http://www.cs.wisc.edu/vdt/>

- **EGEE is running the largest multi-VO grid in the world!**
- **Creating the “grid layer” in e-Infrastructure for research, public service and industry**
- **Key concepts for EGEE**
 - Sustainability – planning for the long-term
 - Production quality
 - And...

- **EGEE is running the largest multi-VO grid in the world!**
- **Creating the “grid layer” in e-Infrastructure for research, public service and industry**
- **Key concepts for EGEE**
 - Sustainability – planning for the long-term
 - Production quality
 - And...
- **Grids are fundamentally about people**
- **... how people in different organisations commit to cooperate**
- **... and how that cooperation can be enabled by operations, training, support, and (most transient of all?) middleware**

Knowledge Utilities as Basis for global Sciences

H. F. Hoffmann/CERN; Nov. 2006



Knowledge Utilities as Basis for global Sciences

H. F. Hoffmann/CERN; Nov. 2006

Networked sciences at the state of the art: e-sciences

Sciences of the 21st century

Openly accessible, user friendly knowledge repositories
based on common efforts across sciences

Lifelong education services



Knowledge Utilities as Basis for global Sciences

H. F. Hoffmann/CERN; Nov. 2006

Networked sciences at the state of the art: e-sciences

Sciences of the 21st century

Openly accessible, user friendly knowledge repositories
based on common efforts across sciences

Lifelong education services

Today's students of the e-sciences will be the innovators and
actors of the 21st century

Knowledge Utilities as Basis for global Sciences

H. F. Hoffmann/CERN; Nov. 2006

Networked sciences at the state of the art: e-sciences

Sciences of the 21st century

Openly accessible, user friendly knowledge repositories
based on common efforts across sciences

Lifelong education services

The emergence of a networked knowledge society in the next twenty to thirty years is a major paradigm shift from the industrial model of the 19th and 20th century. This transition is of crucial importance in opening up new opportunities for education, social inclusion, and more efficient use of resources.

Information and communication technologies are the effective tools of this transition. © 8/2003 The Club of Rome

Today's students of the e-sciences will be the innovators and actors of the 21st century