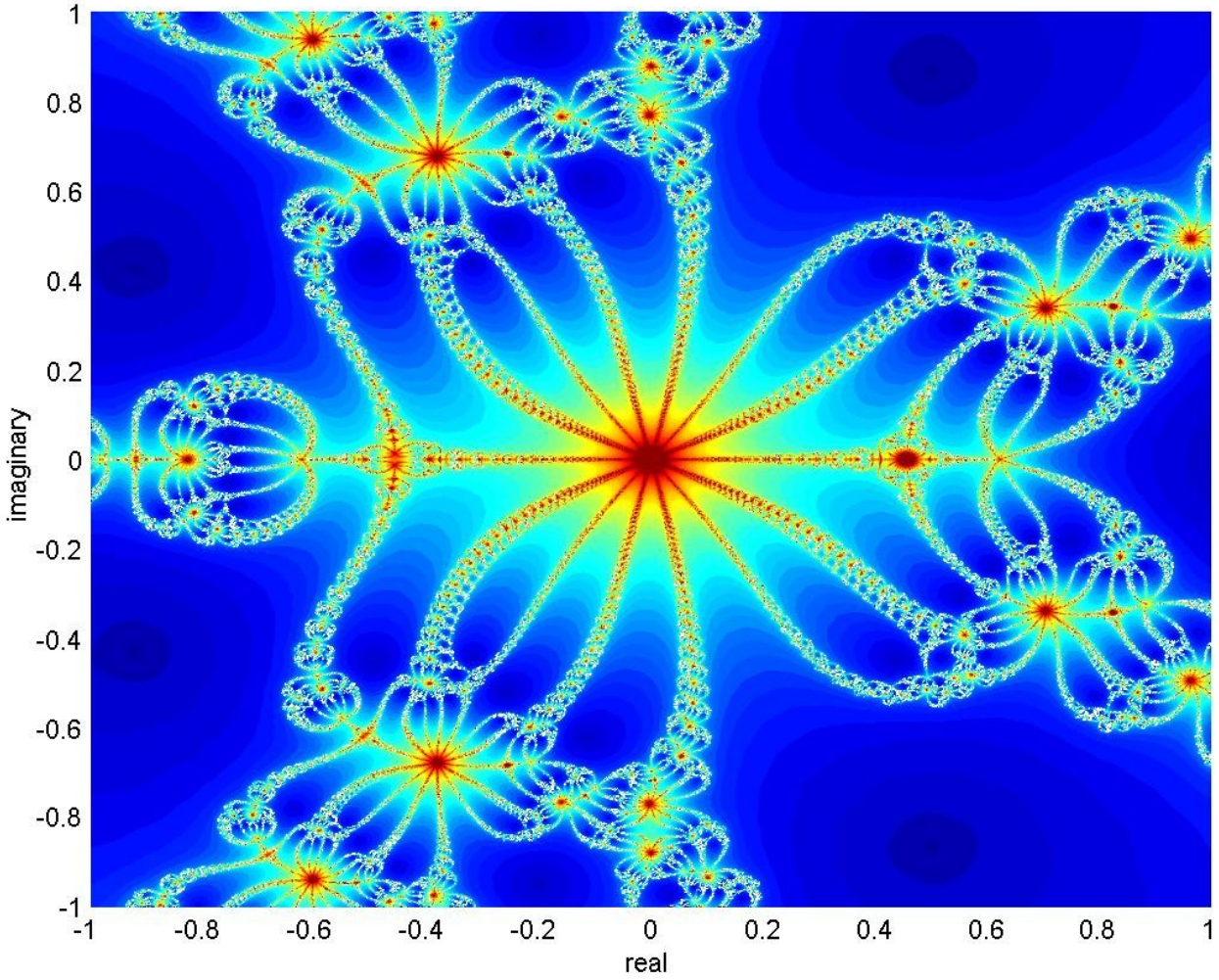
Enabling Grids for E-science in Europe

Nicosia, April 23rd, 2007

Scientific Areas and Existing Virtual Organizations

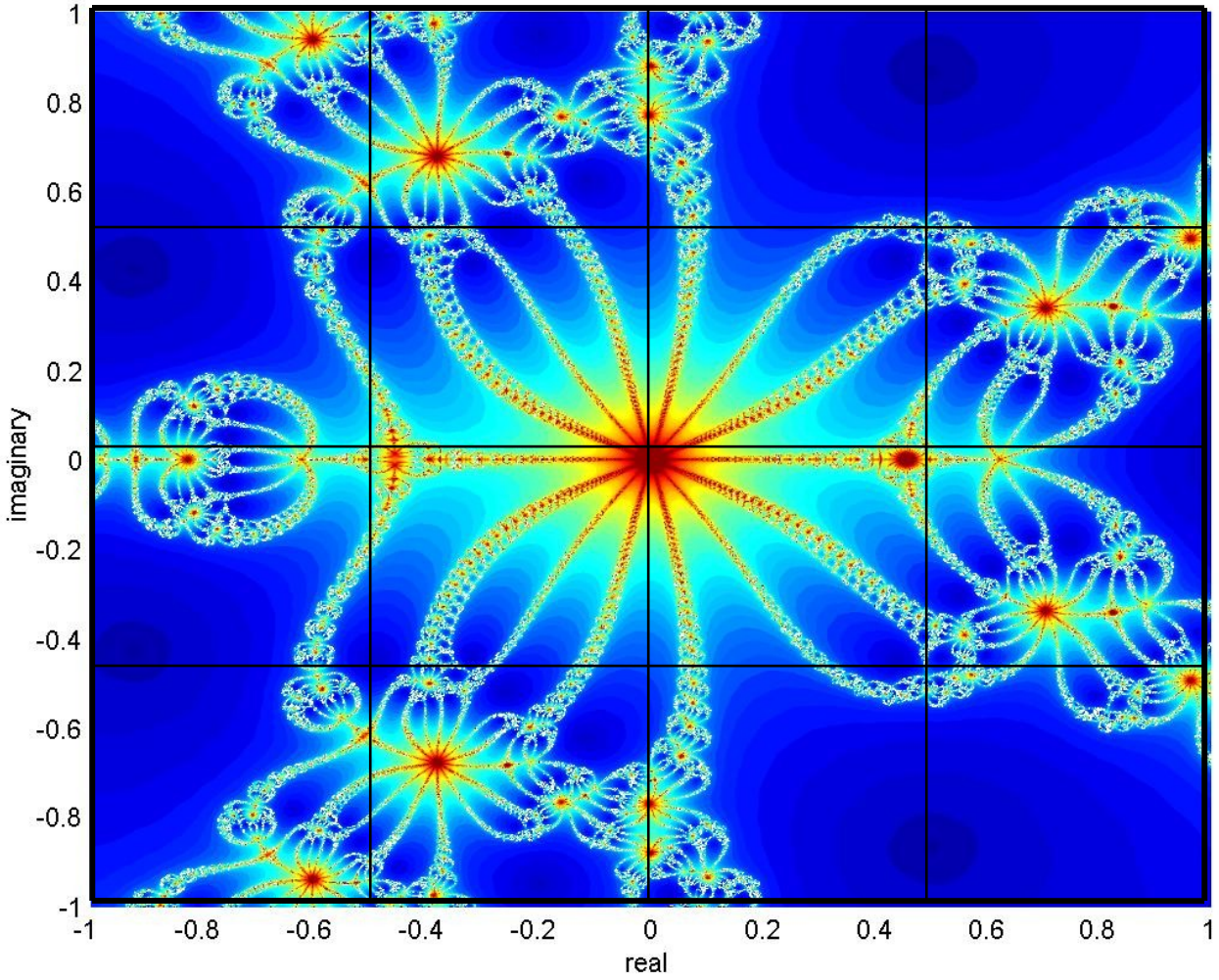
Fotis Georgatos <fotis@mail.cern.ch>
Trainer, University of Cyprus

An application at a single computer



An application on the Grid

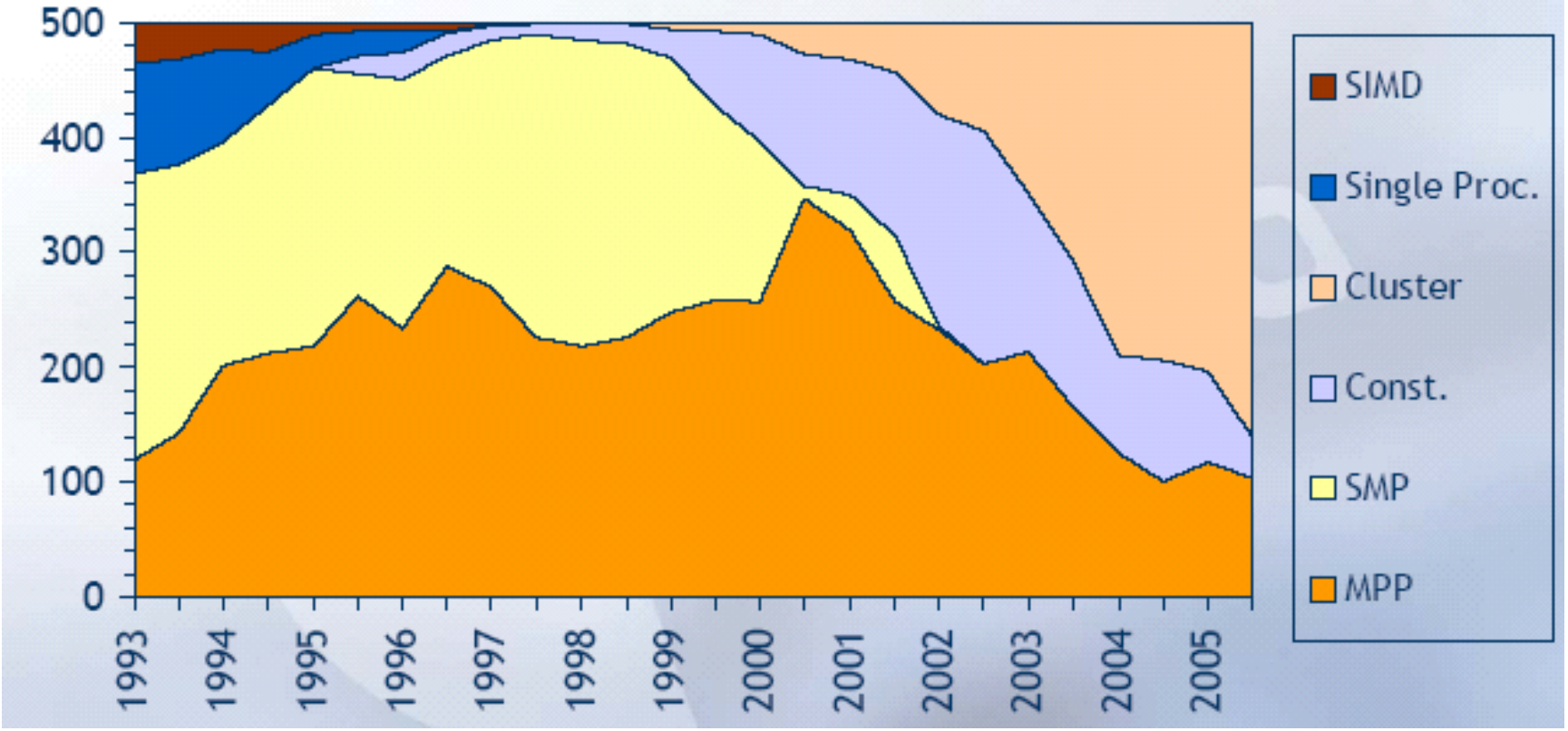
Enabling Grids for E-science in Europe



What evolutions make Grid emerge

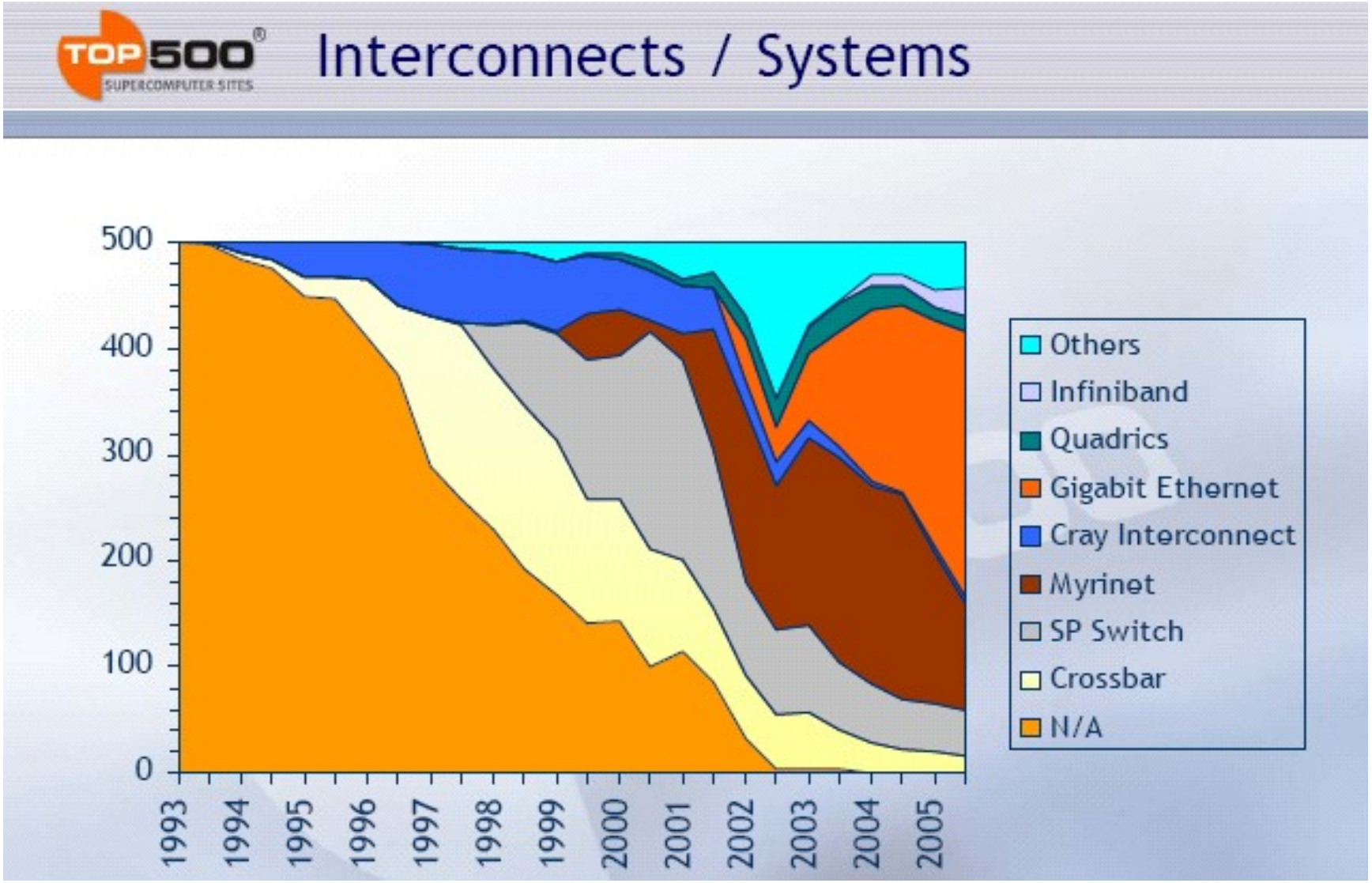
Enabling Grids for E-science in Europe

TOP500 SUPERCOMPUTER SITES — Architectures / Systems

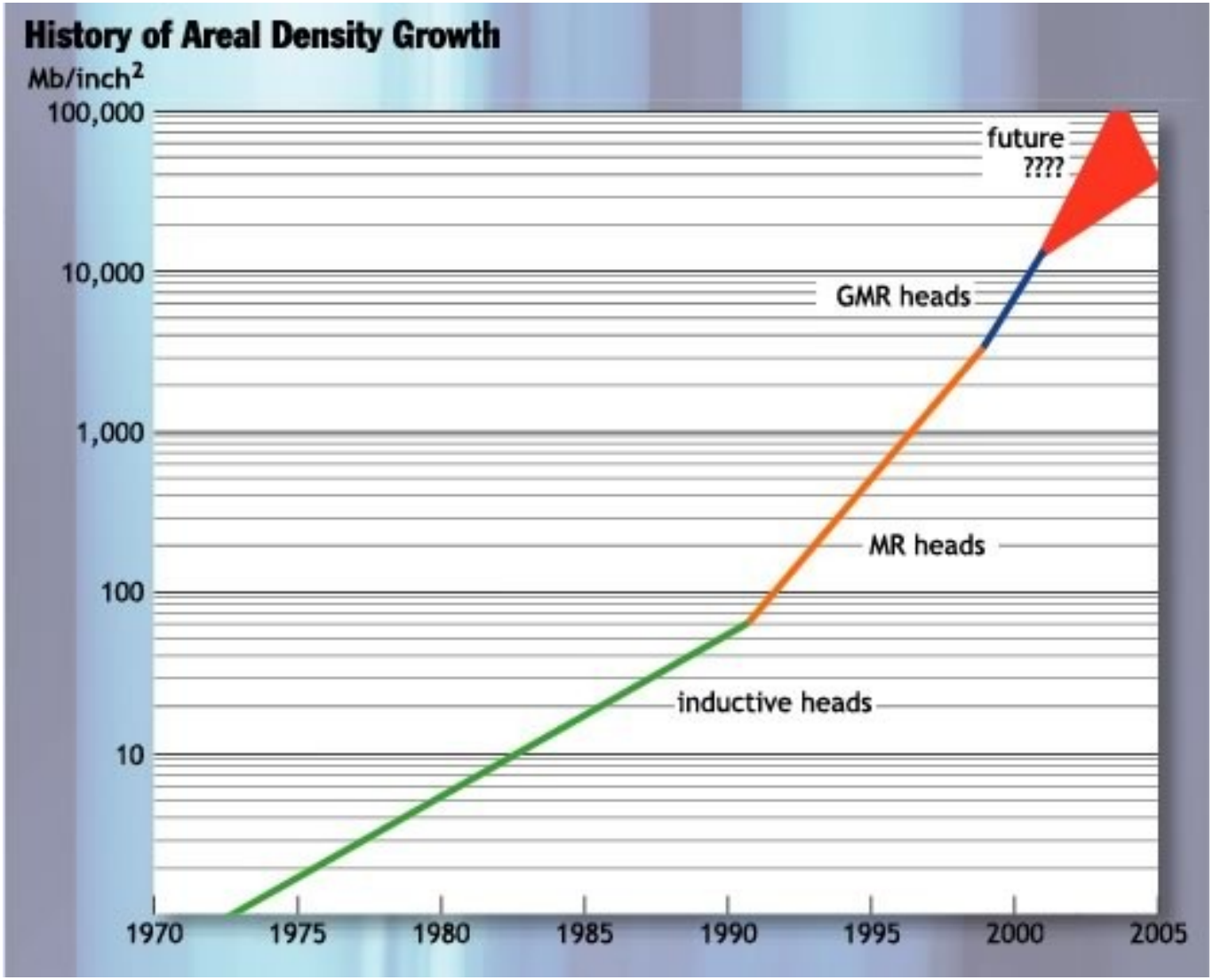


What evolutions make Grid emerge

Enabling Grids for E-science in Europe

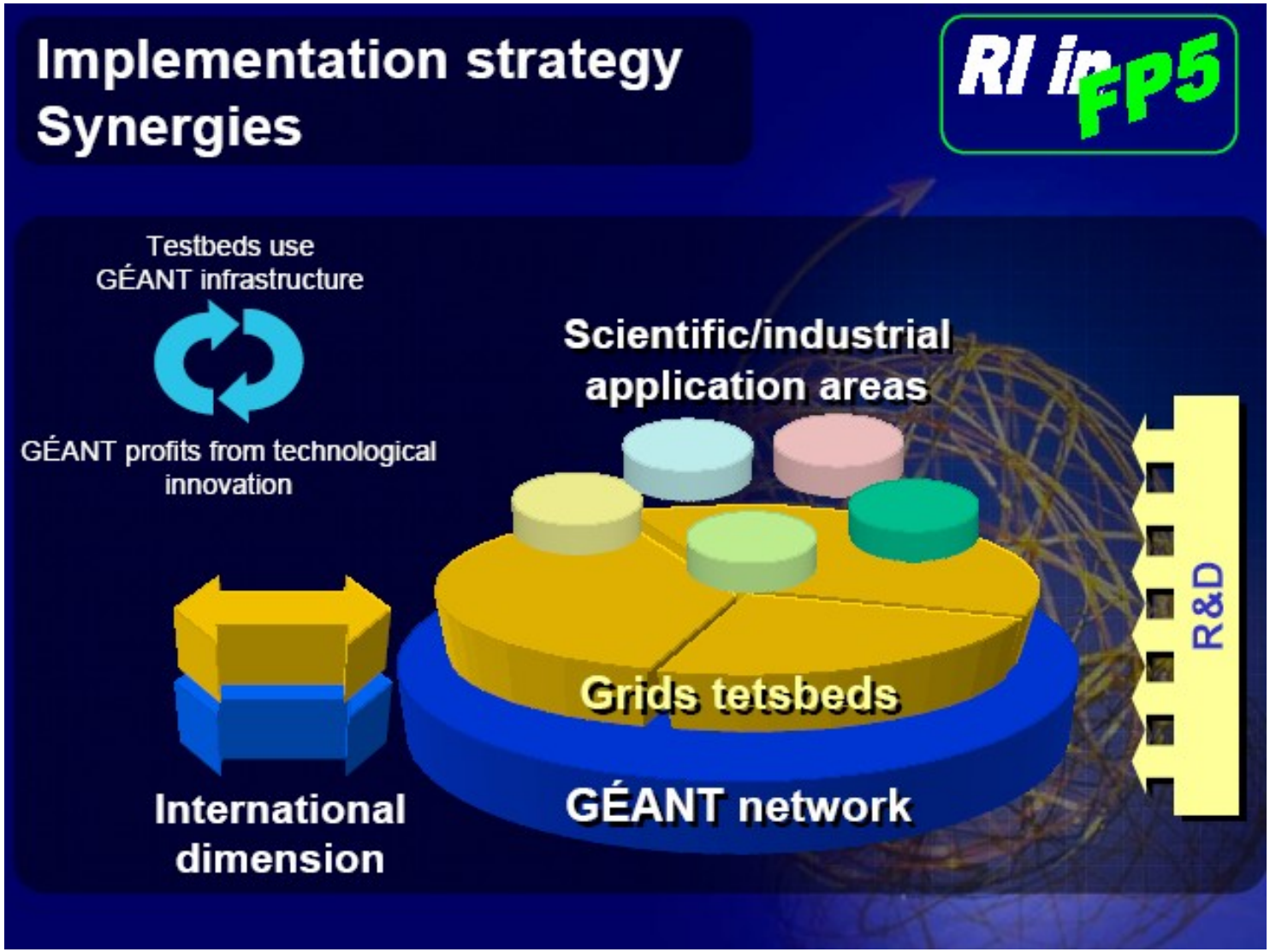


What evolutions make Grid emerge



Why does Europe need the Grid

Enabling Grids for E-science in Europe

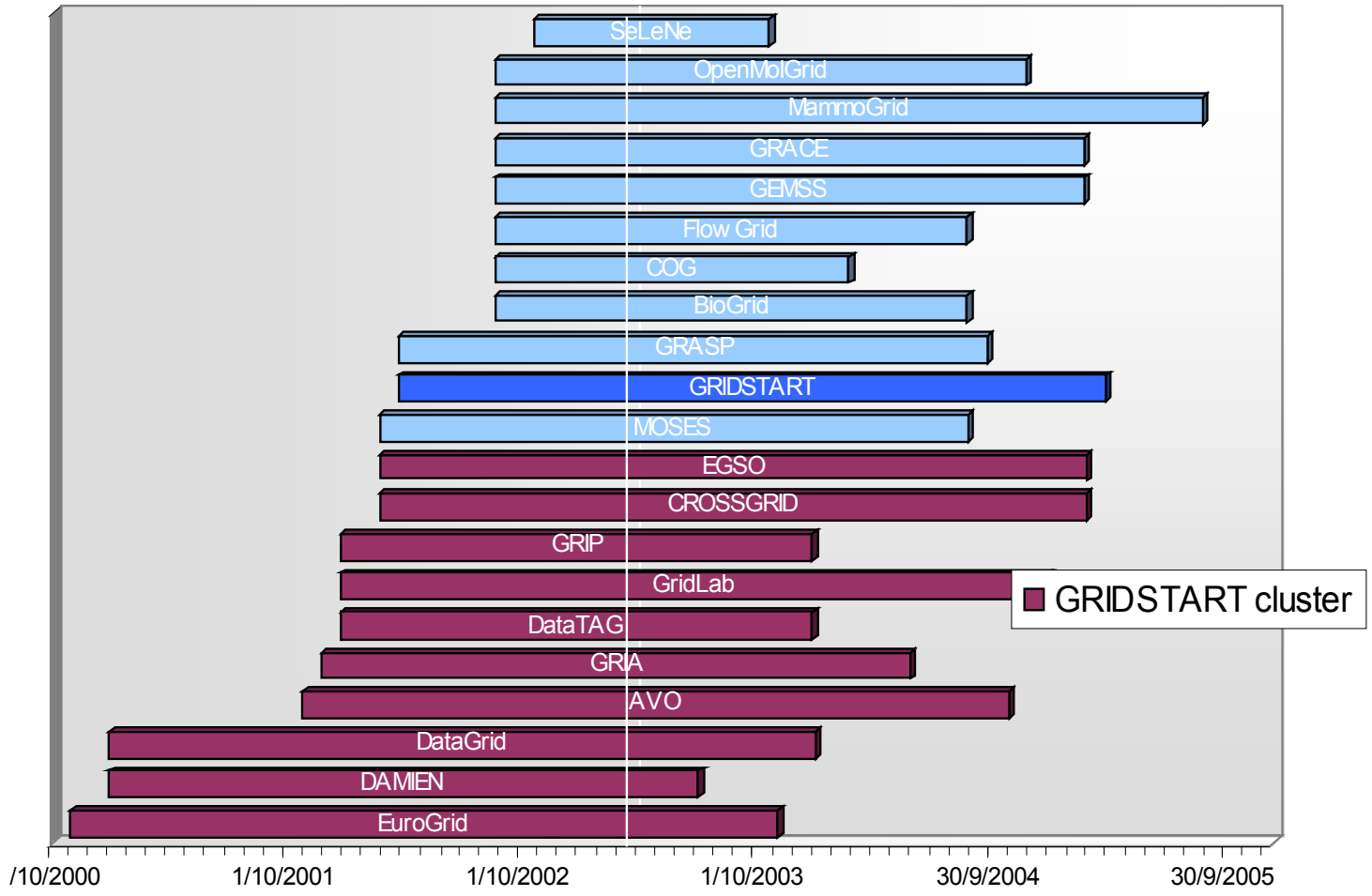


Why NRENs need the Grid

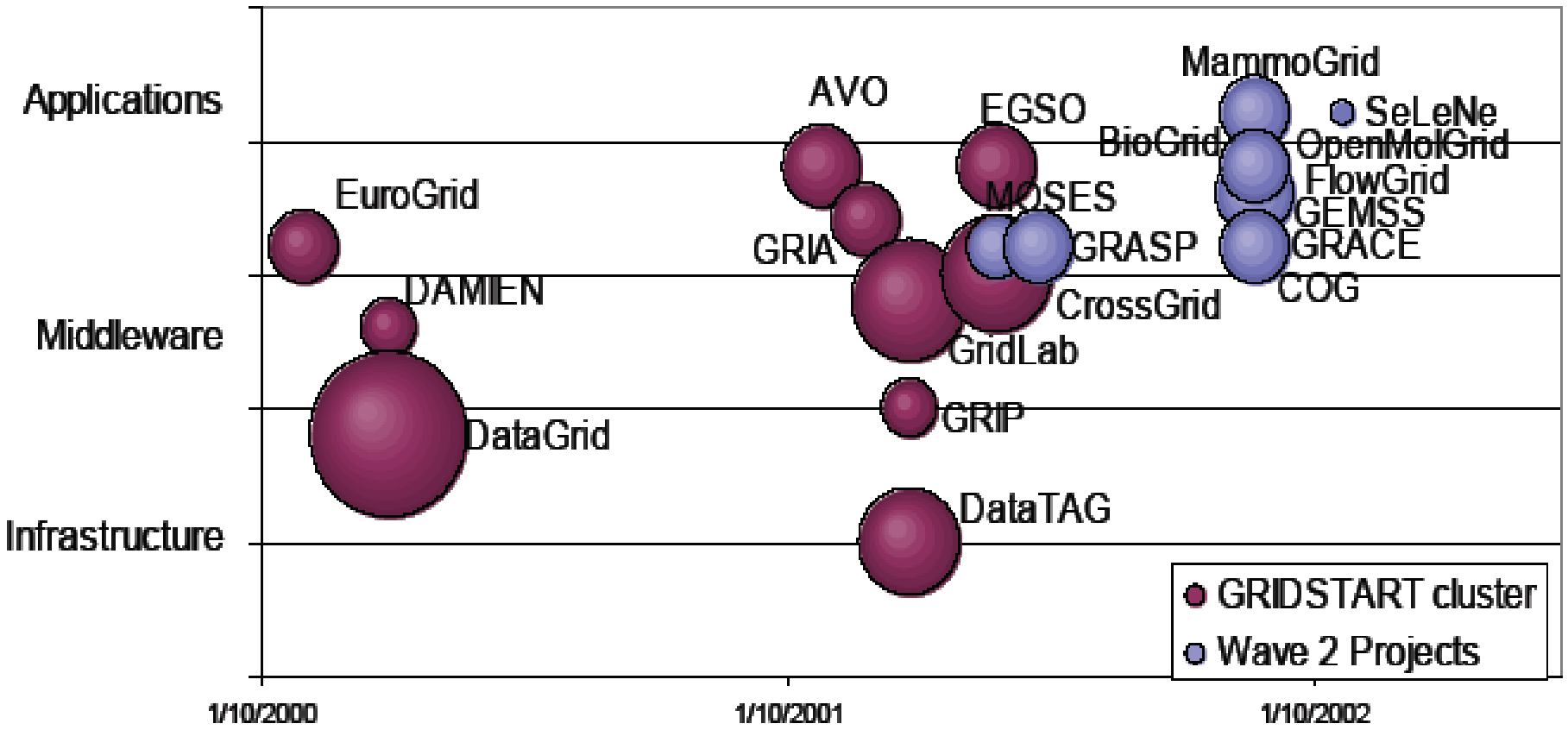
Enabling Grids for E-science in Europe

First and second wave of projects

Enabling Grids for E-science in Europe



First and second wave of projects



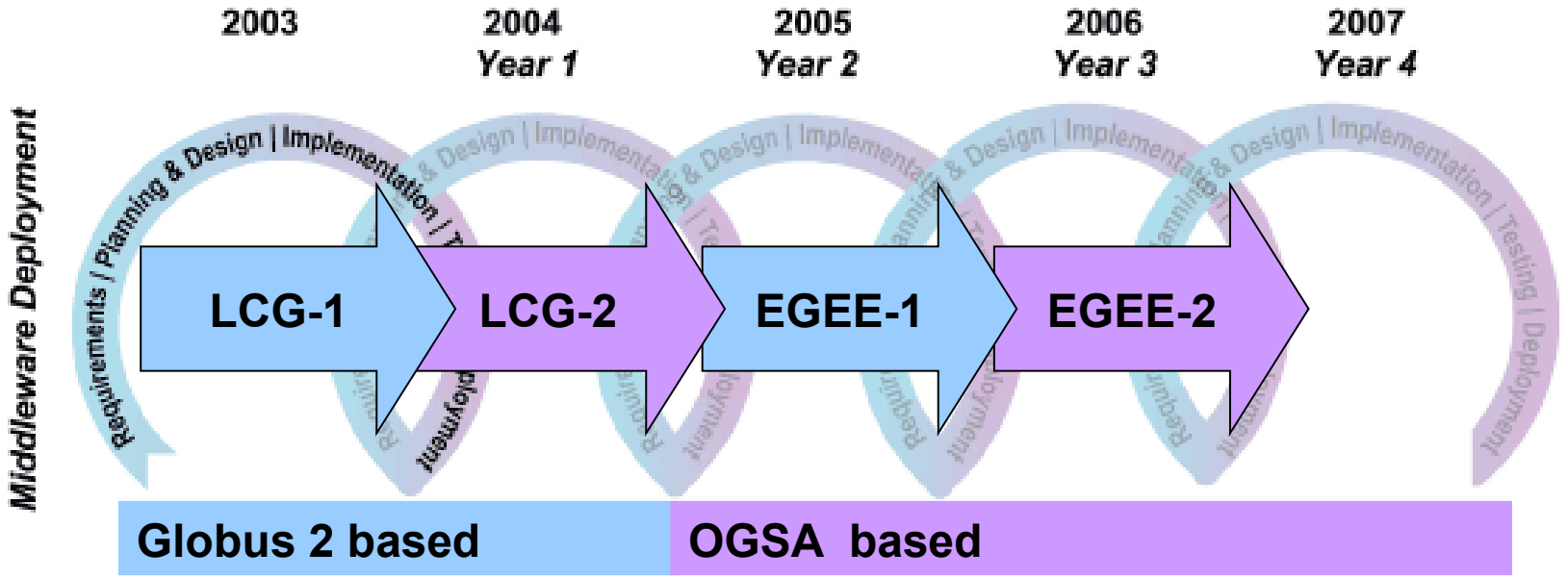
The birth of EGEE

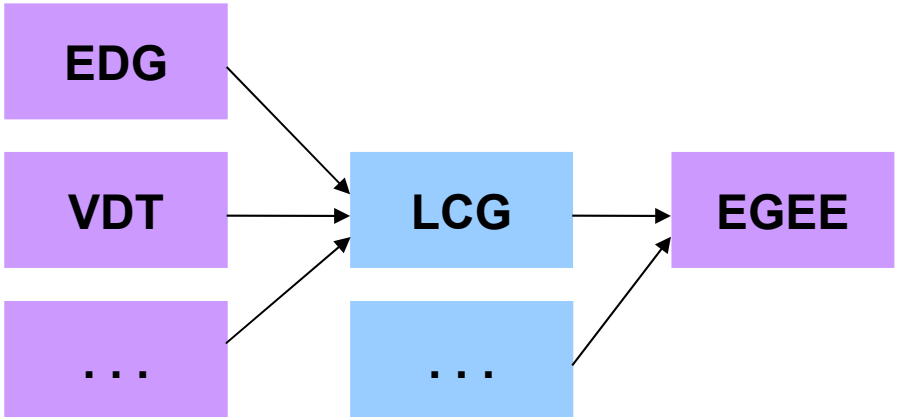
Enabling Grids for E-science in Europe

- EU and EU member states major investment in Grid Technology
- Several good prototype results
- Next Step:
  - Leverage current and planned national programmes
  - work closely with relevant industrial Grid developers and NRNs
  - build on existing middeware and expertise
  - create a general European Grid production quality infrastructure
  - This can be achieved for a minimum of €100m/4 years on top of the national and regional initiatives

LCG and EGEE

Enabling Grids for E-science in Europe





The EGEE vision

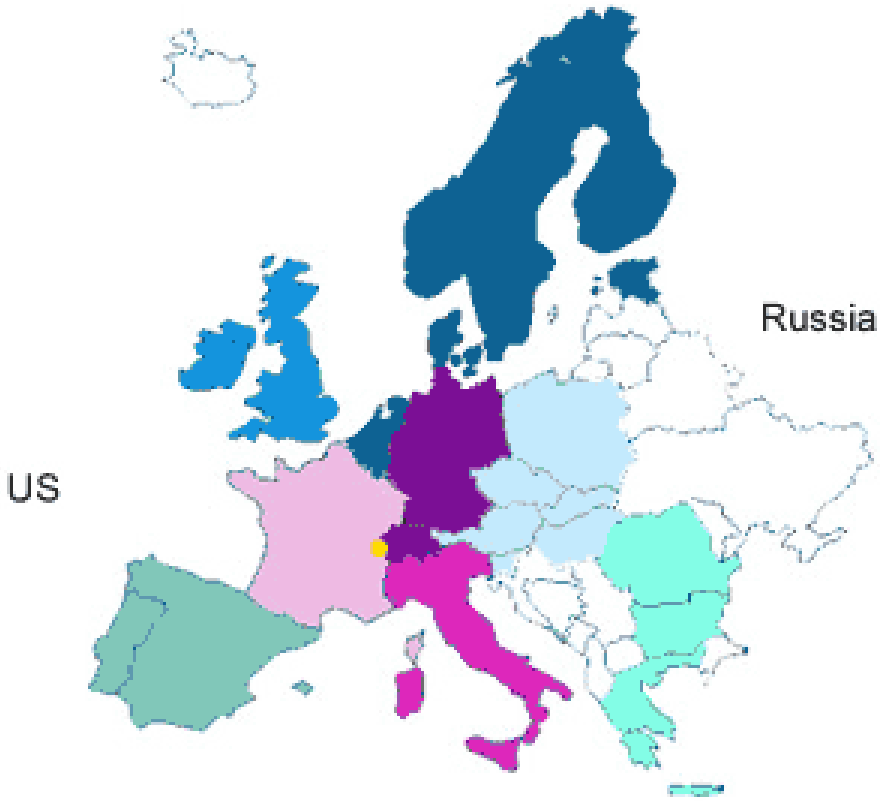
Enabling Grids for E-science in Europe

- Creation of a wide Paneuropean Grid infrastructure, incorporating current and future Science Research Networks
- Provide for the distributed european research communities 24/7 access to computational resources, regardless of geography
- Emphasis on the User of Grid technologies, rather than Development
- Support of multiple application fields, by a large scale infrastructure that can integrate and consolidate any further deployed resources
- Provision of education and support to end users

Which people cooperate for EGEE

Enabling Grids for E-science in Europe

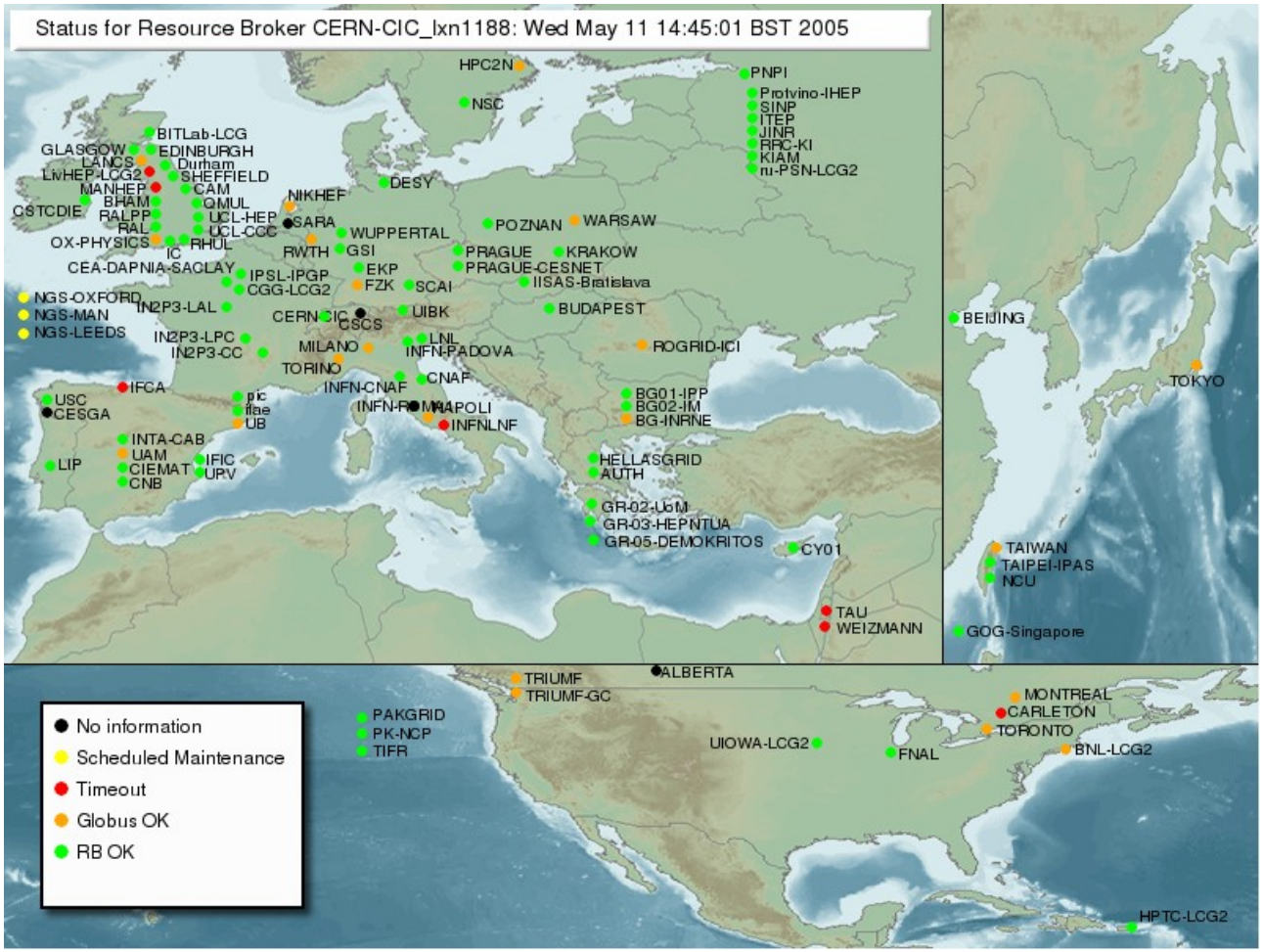
- >100 leading institutes in >40 countries, organized according to regions
- Provision of national networks, aiming at European cooperation



- CERN
- Central Europe (Austria, Czech Republic, Hungary, Poland, Slovakia, Slovenia)
- France
- Germany and Switzerland
- Ireland and UK
- Italy
- Northern Europe (Belgium, Denmark, Estonia, Finland, The Netherlands, Norway, Sweden)
- Russia
- South-East Europe (Bulgaria, Cyprus, Greece, Israel, Romania)
- South-West Europe (Portugal, Spain)

Where is the EGEE infrastructure

Enabling Grids for E-science in Europe



New map: http://goc03.grid-support.ac.uk/googlemaps/lcg2.html

Grid Middleware

Enabling Grids for E-science in Europe

- Operating System:
  - Linux(+GNU), usually a RHEL3-like, fi. Scientific Linux 3.0.7, Fedora Core 3 κλπ.
- Middleware:
  - gLite v3.0.5 (was until recently: LCG v2.7.0)
- Libraries & Applications:
  - Any software that system administrators of the infrastructures have installed (it is though possible for a user to install his own programs during a job execution)

The architecture of LCG/EGEE

Enabling Grids for E-science in Europe

- LCG stands for LHC Computing Grid, which a CERN's project
- EGEE is a collection of distributed resources, geographically dispersed
- LCG/EGEE Users:
  - Are Organized according to the concept of **Virtual Organizations, VOs**
  - They run applications, ignoring:
    - Where a process runs
    - Where input data comes from
    - Where output data goes to
- LCG/EGEE software consists of:
  - *Workload Management System*
  - *Data Management System*
  - *An Information System*
  - *An Authorisation and Authentication System*
  - *An Accounting System* (RGMA)
  - *Various monitoring services*
  - *Various installation services*

Where current software comes from

Enabling Grids for E-science in Europe

Component	LCG	EGEE	EDG	EDT	INFN-GRID	Globus	Condor	Other
Basic middleware								
Globus 2.4.3						√		
ClassAds 0.9.4							√	
Security								
MyProxy								√
VO management								
LDAP-based			√					
VOMS	√	√	√					
Workload management								
Condor/Condor-G 6.6.5							√	
EDG WMS	√		√					
Data management								
Replica Manager	√		√					
Replica Location Service			√			√		
LCG File Catalog	√						√	
Disk Pool Manager	√							
GFAL	√							
LCG DM tools	√							
Fabric management								
LCFG	√		√					√
Quattor	√		√					
YAIM	√							
LCAS/LCMAPS			√					
Monitoring								
GridICE					√			
Information system								
MDS						√		
Glue Schema				√				√
BDII	√							
R-GMA		√	√					
LCG Information tools	√							

- EDG
- LCG
- EGEE
- INFN
- Globus
- Condor
- Other (EDT, VDT, etc)

Sciences and Grid

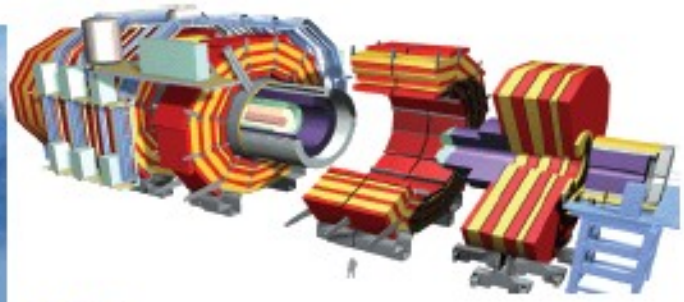
Enabling Grids for E-science in Europe

- Physics and Astronomy
  - High Energy Physics, Radioastronomy
- Bioinformatics
  - Study of Human Genome in favor of understanding genetic diseases, Protein synthesis
- Medicine and Public Health
  - Medical data visualization, diagnosis and cure, Pharmaceutics
- Natural Resources and the Environment
  - Weather forecasting, Geosciences and seismology, modeling and forecasting of complex systems, fi. ocean currents, air mass flow etc
- Engineering and Applied Sciences
  - Buildings and Civil Engineering, Economy and Industry, Data mining
- Computational Chemistry, Material Sciences, Nanotechnology
  - Design of new materials and study from molecular level up

Large Hadron Collider @ CERN

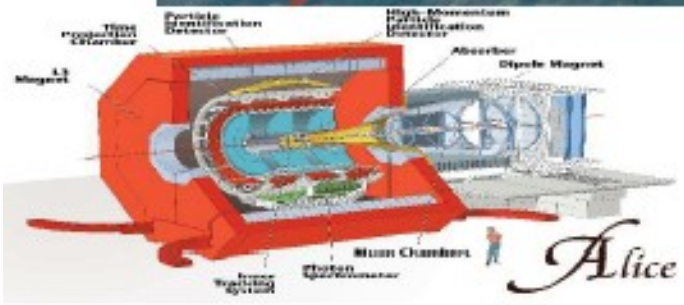
Enabling Grids for E-science in Europe

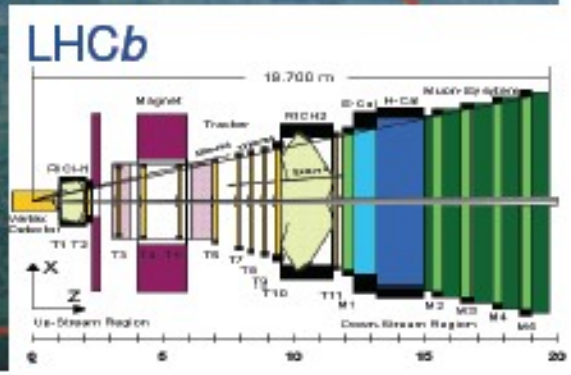
Atlas



CMS

LHC





Which are the Virtual Organizations

Enabling Grids for E-science in Europe

- VOs affiliated to LHC/CERN
  - ALICE VO
  - ATLAS VO
  - CMS VO
  - Geant4 VO
  - LHCb VO
  - SixTrack VO
- Other VOs related to HEP
  - Babar VO
  - D0 VO
  - H1 VO
  - ILC VO
  - PhenoGrid VO
  - Planck VO
  - Zeus VO
- VOs of other sciences
  - Biomed VO
  - CompChem VO
  - EGEODE VO
  - ESR VO
  - E-earth VO
  - Magic VO
- VOs of regional interest
  - SEE VO
  - HellasGrid VO
  - HellasGrid-Demo VO
  - INFN VO
  - DutchGrid VO
  - Desy VO
  - CESGA, SWETEST, IFIC, etc

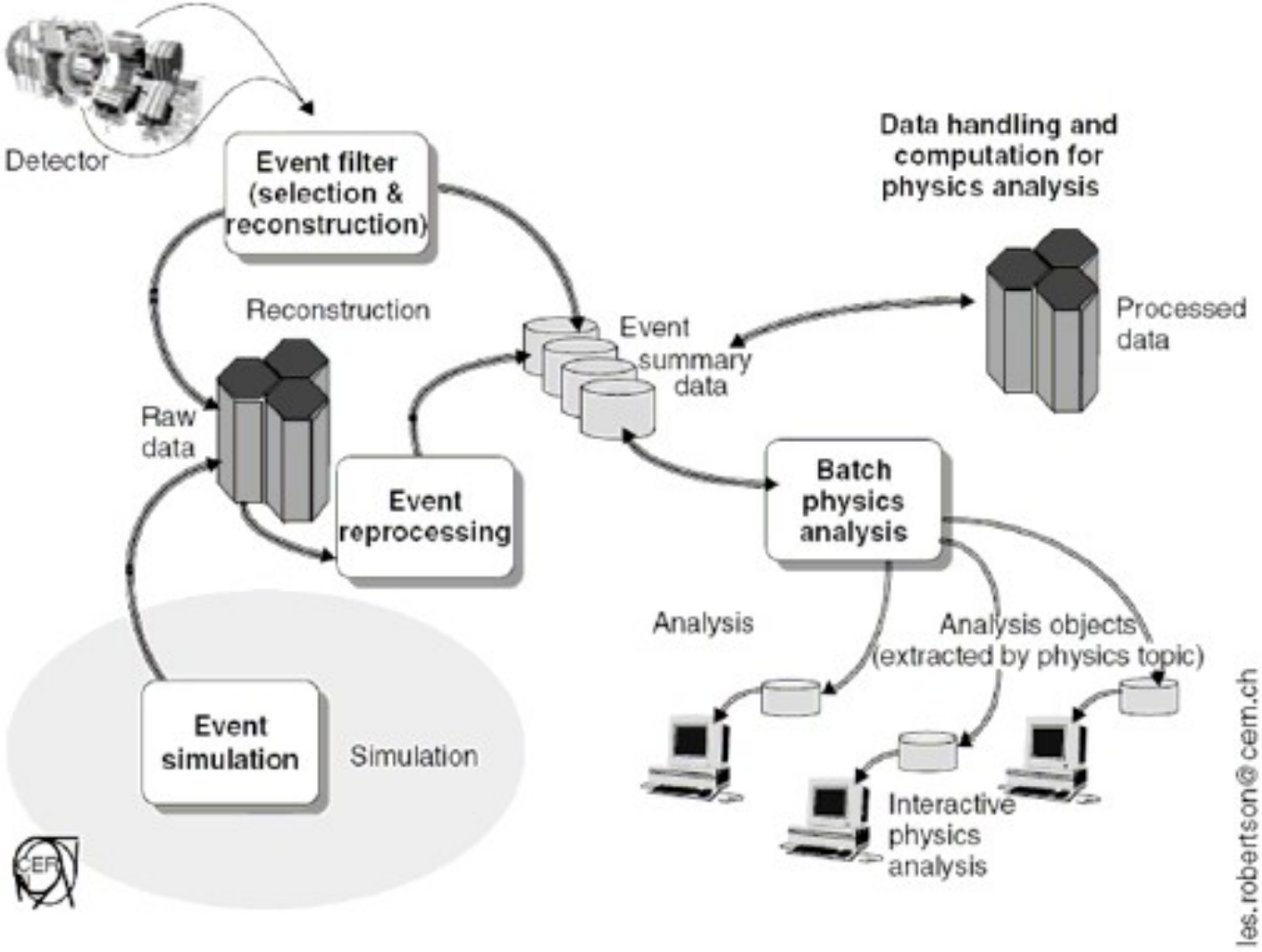
What software do VOs «run»

Enabling Grids for E-science in Europe

Each VO can install or demand special software, which covers its specialized needs:

- ATLAS: atlas software (big collection)
- CMS: cmkin, cobra, famos, geometry, ignominy, orca, oscar
- ALICE: alien, alice, root, proof
- LHCb: dirac, boole, DC, decfiles, gauss, paramfiles
- BIOMED: gate, cdss, gps@, gromacs, simri3d, gptm3d
- ESR: (earth science specific…)

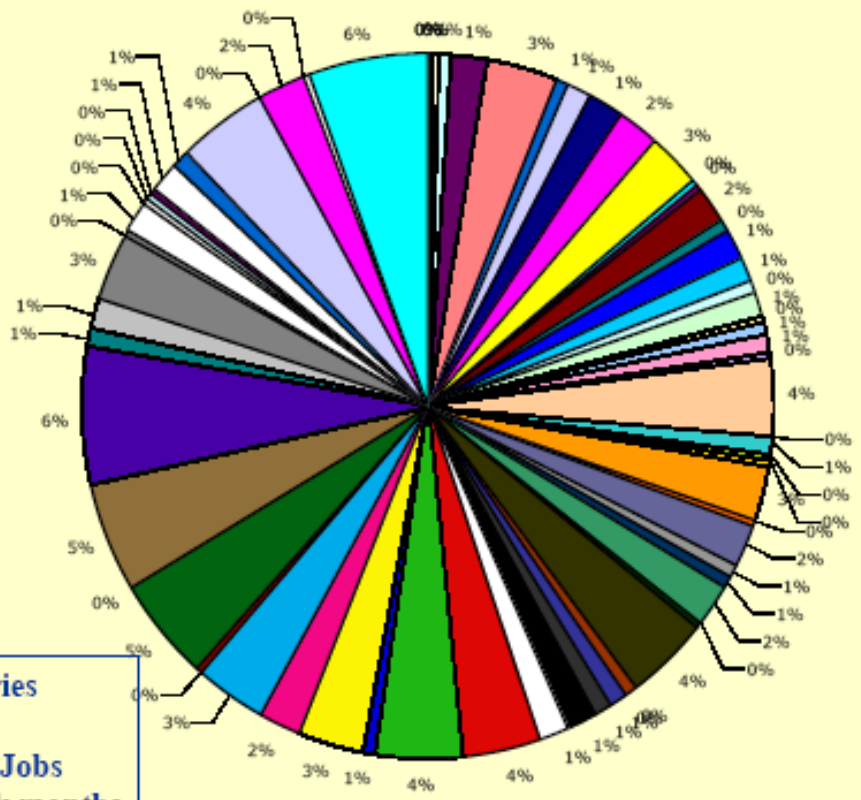
The principles of CERN VOs



An example from an ATLAS run

Enabling Grids for E-science in Europe

Fraction of GRID jobs per institute



20 countries
69 sites
~ 260000 Jobs
~ 2 MSi2k.months

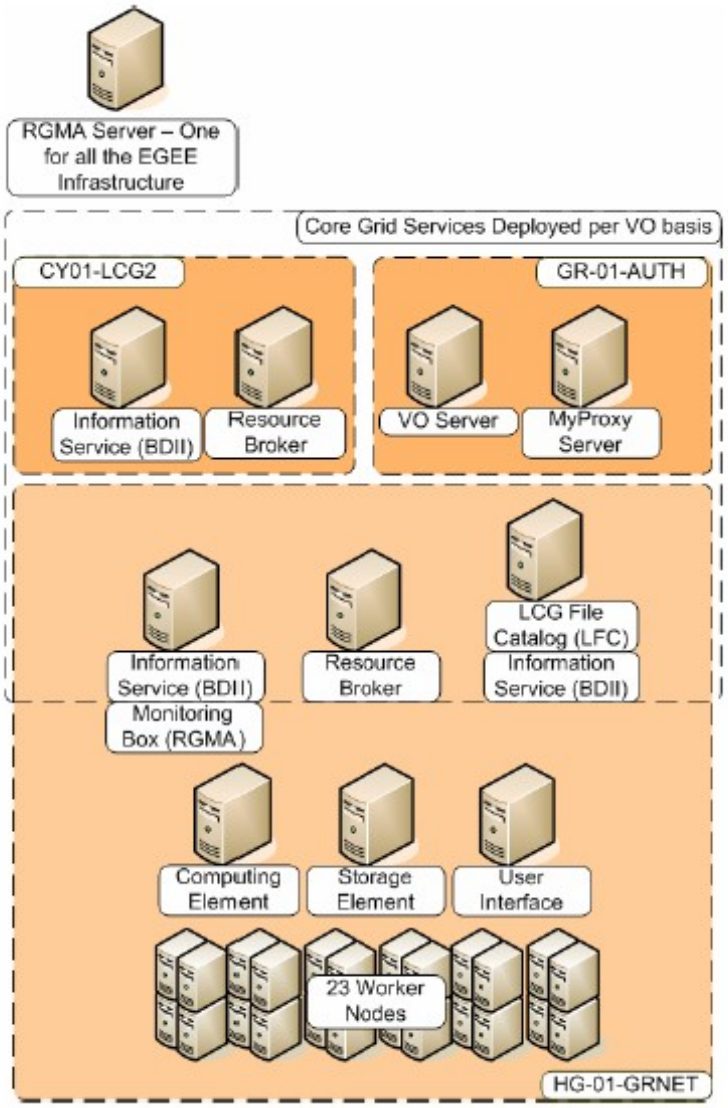
June 05 ATLAS Week - M. Cobal

Requirements of LHC/CERN VOs

Enabling Grids for E-science in Europe

	ALICE	ATLAS	CMS	LHCb
SE GB/ cpu	30	20	50	-
WN Disk GB/job	2.5	2	1	5
WN memory MB/job	600	300 (1 GB for pileup at selected sites)	500	500
Longest job (@ - 2 GHz)	8 h	24 h	72 h (1 week for Oscar)	24h
SW installation space (GB)	0.5 GB in shared area	15 GB	0.7 GB (production) 20 GB (analysis) in shared area	0.5 GB

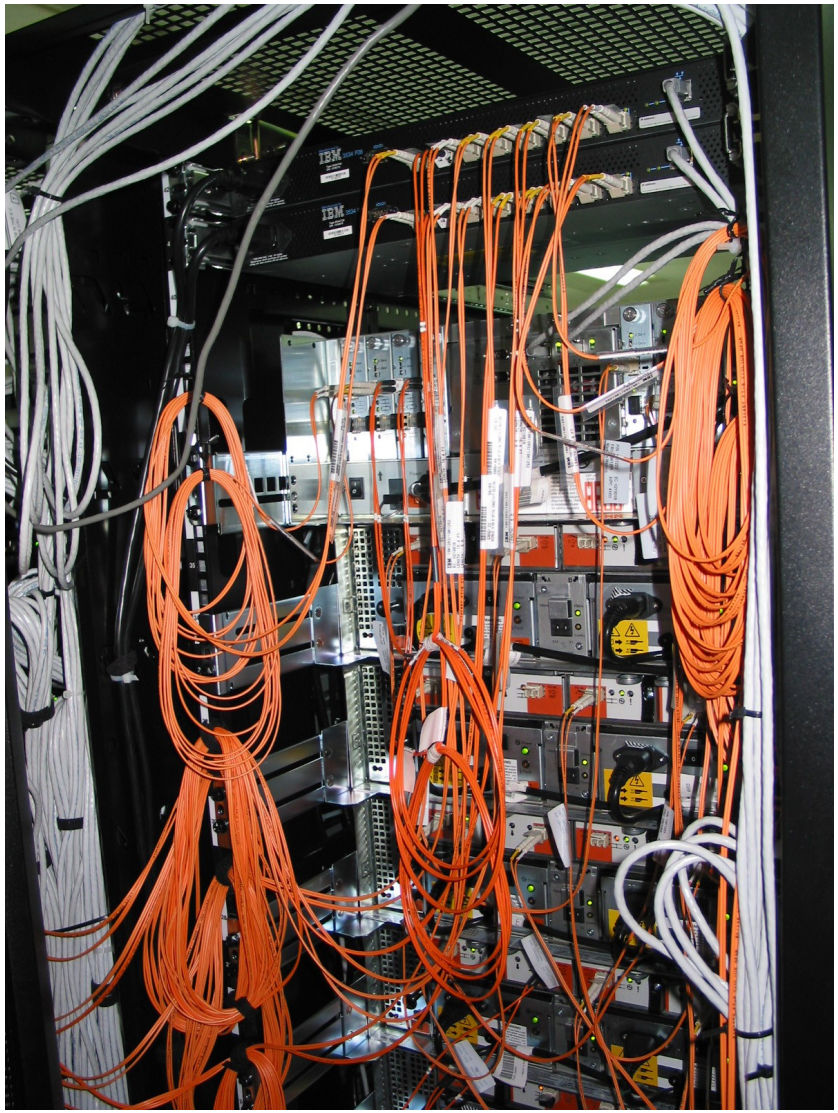
Dissecting a VO: SEE, HellasGrid



- User directory:
  - VO server & Myproxy
- Resources directory:
  - BDII (LDAP based!)
- Computational Resources:
  - Resource Broker (RB)
- Storage Resources:
  - LCG File Catalog (LFC)
- Local infrastructures:
  - CE & WNs, SE, UI κλπ.

HellasGrid I infrastructure, Isabella

Enabling Grids for E-science in Europe

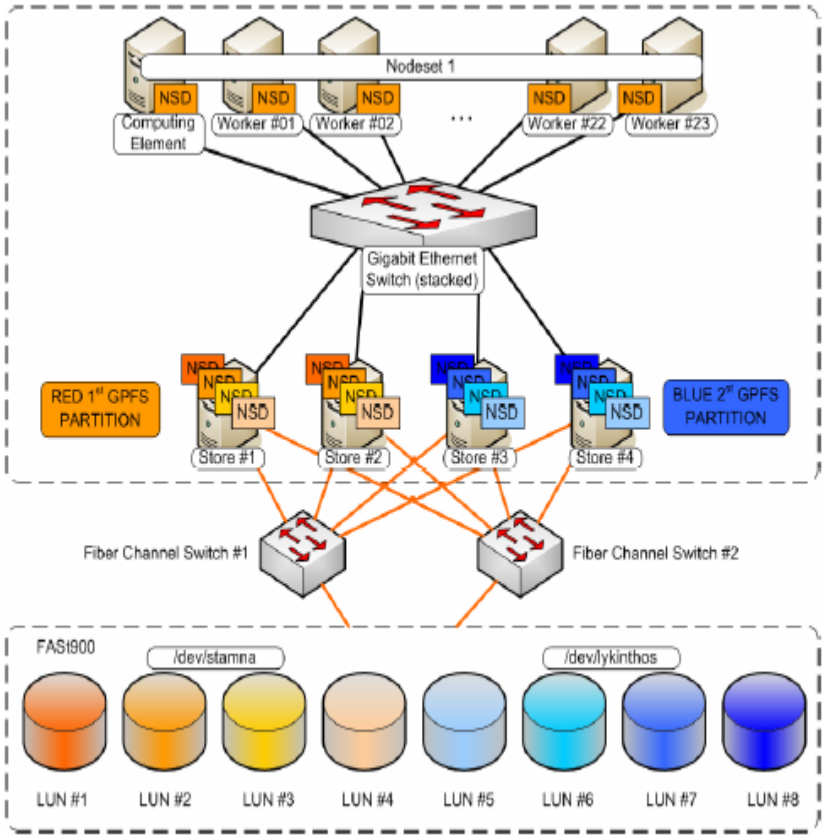


HellasGrid project, Phases I & II

Enabling Grids for E-science in Europe

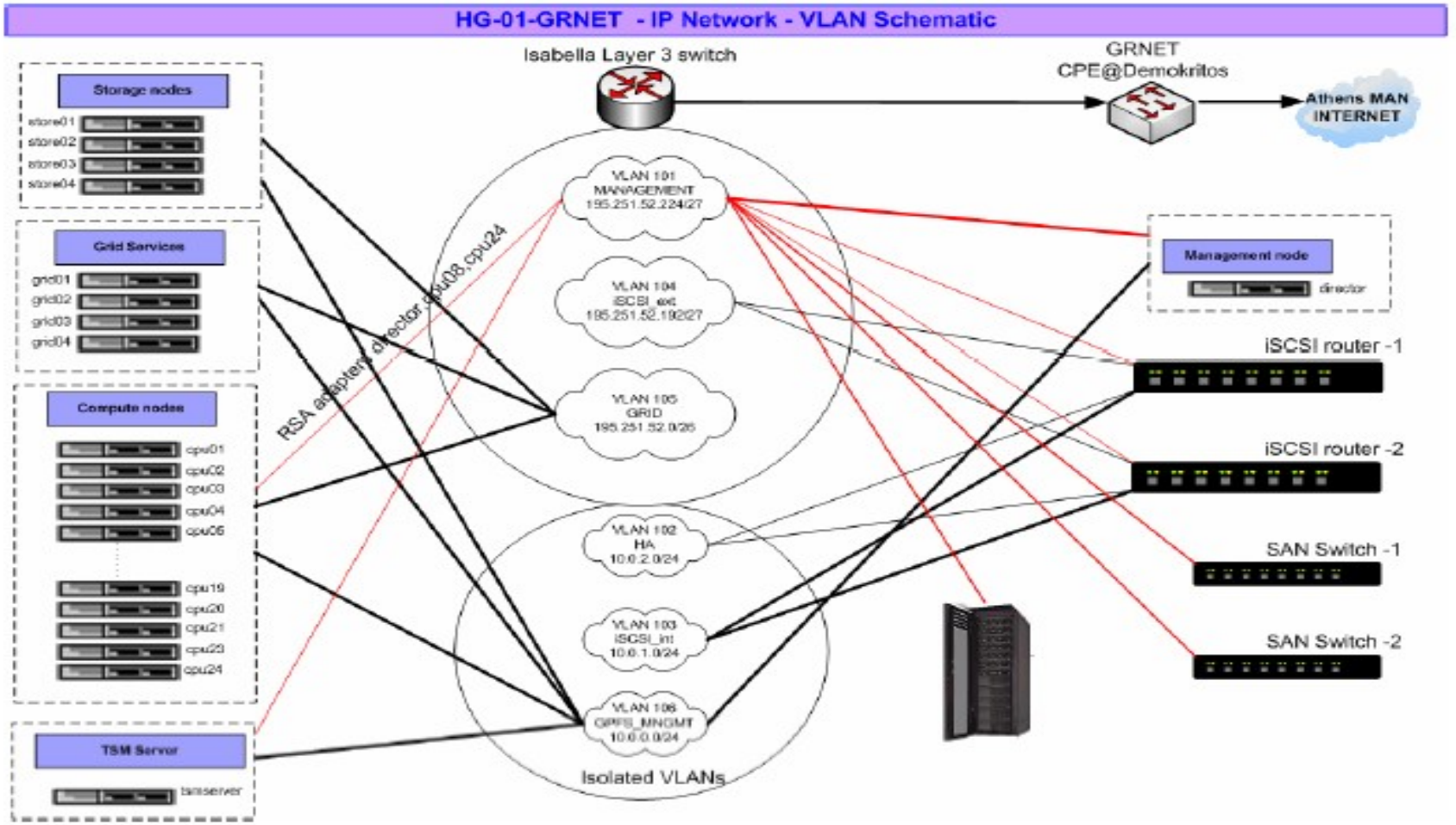
- HellasGrid I
  - Located at Demokritos, Agia Paraskevi, Athens (aka. Isabella)
  - 34 dual Intel Pentium Xeon @ 2.8GHz, 1GB RAM, 140GB HDD, 2x Gigabit
  - IBM FAStT900 Storage Area Network, integrated system
    - Redundant Fibre Channel Controllers with 1Gbyte Cache
    - 70x146.8GB= **10,276TB raw storage capability**
    - Fully automated solution, **hot spare + hot swap**
  - Tape Library with a capacity up to ~30 TBytes
  - Delivered to ΕΔΕΤ by IBM during December 2004
- HellasGrid II
  - 5 more physical nodes: ΕΚΤ, ΙΕΣΕ, ΑΠΘ, ITE, ITY
  - ~700 CPUs x86_64, 2 GB RAM, 80GB HDD, 2x Gigabit
  - ~20 TBytes total storage capacity provided by SAN solutions
  - ~50TBytes Tape Library
  - Under installation (equipment has been already delivered)

HellasGrid I infrastructure, Isabella



- The first node of the HellasGrid infrastructure has been a great tool for building a knowledge base.
- The experience with it is going to be exploited during the second phase of the project, in benefit of the newer nodes and users.
- Outstanding and very unconventional organization of the SAN system and its filesystems.

HellasGrid I infrastructure, Isabella



Ready and waiting for your jobs!

Enabling Grids for E-science in Europe

Q & A

Enabling Grids for E-science in Europe