



Enabling Grids for E-scienceE

VO Management in EGEE

Maurice Bouwhuis

SARA computing and networking services

**Joint EGEE and OSG Workshop on VO Management in
Production Grids**

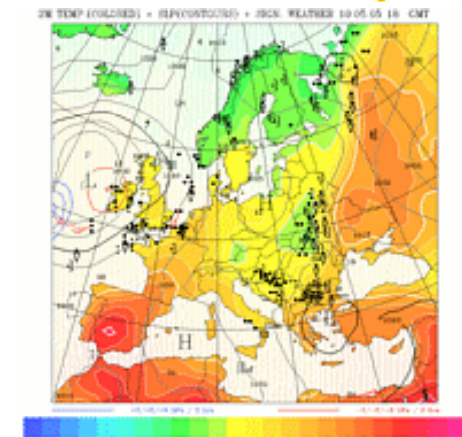
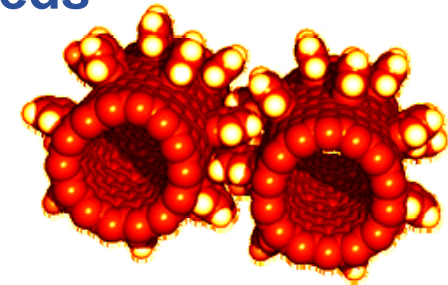
HPDC 2008

Boston, USA

www.eu-egee.org



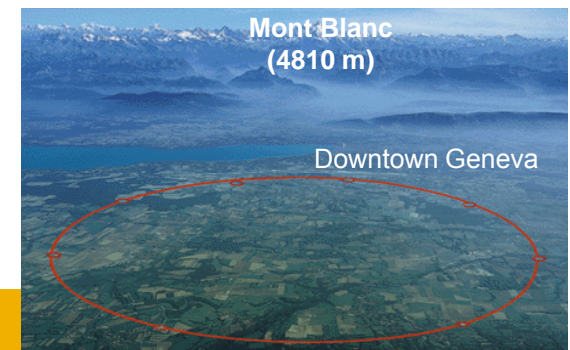
- Science is becoming increasingly **digital**, needs to deal with increasing amounts of data and computational needs
 - **Simulations** get ever more detailed
 - Nanotechnology – design of new materials from the molecular scale
 - Modelling and predicting complex systems (weather forecasting, river floods, earthquake)
 - Decoding the human genome
 - **Experimental Science** uses ever more sophisticated **sensors** to make precise measurements
 - Need high statistics
 - Huge amounts of complex data
 - Serves user communities **around the world**
- **EGEE = enabling grids for e-science**

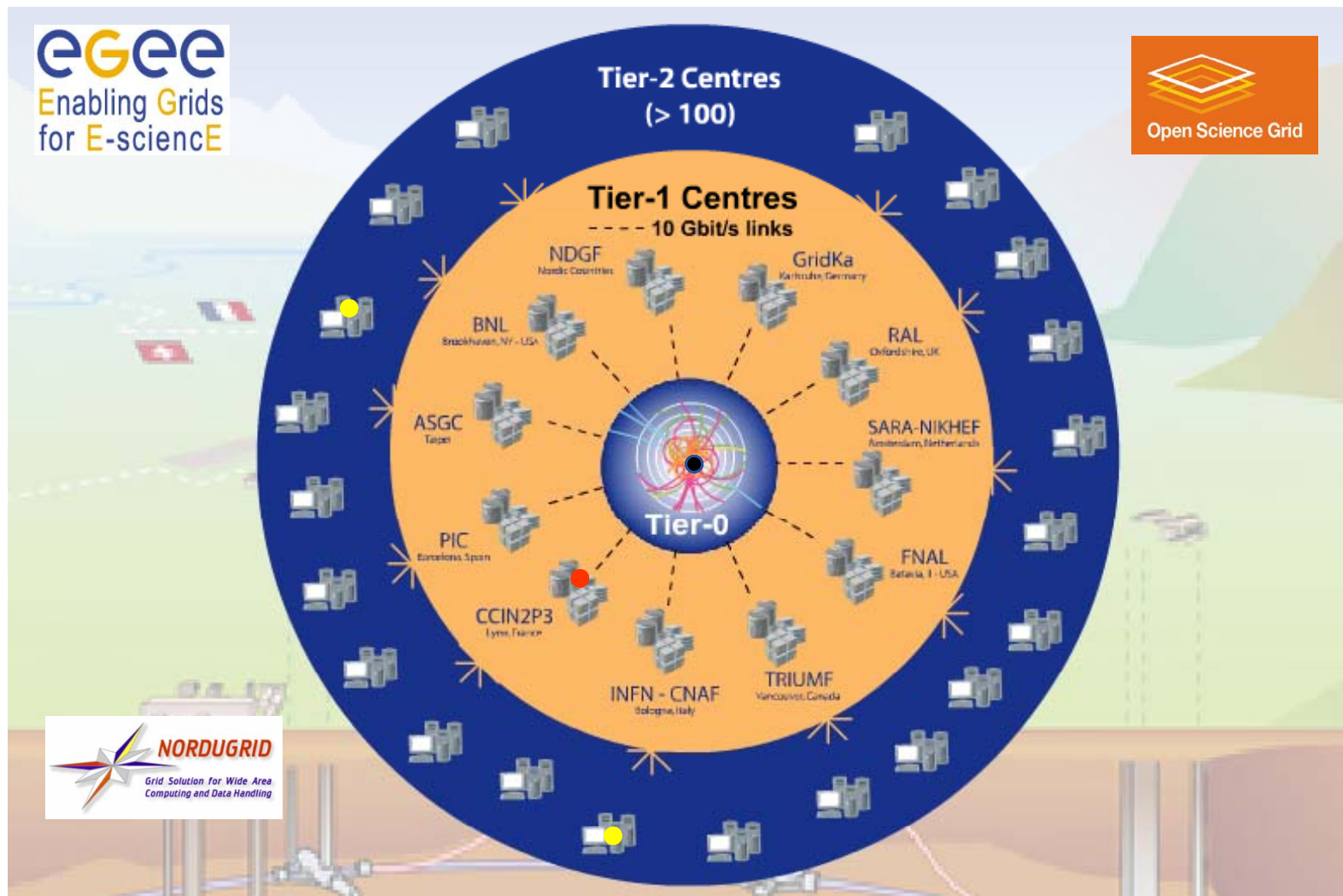


Large Hadron Collider



- 27 km circumference tunnel
- Due to start up in 2008
- 40 Million Particle collisions per second
 - Online filter reduces to a few 100 “good” events per second recorded on disk and magnetic tape at 100-1,000 MegaBytes/sec
 - ~15 PetaBytes per year for all four experiments
- Data analyzed by 100s of research groups world wide





Millions of chemical compounds available in laboratories

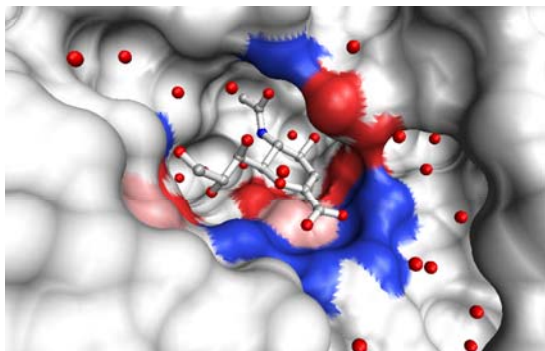


High Throughput Screening
 2\$/compound, nearly impossible

300,000 Chemical compounds:
ZINC &
 Chemical combinatorial library



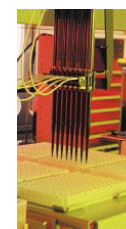
Molecular docking (**Autodock**)
 100s CPU years, TBs data



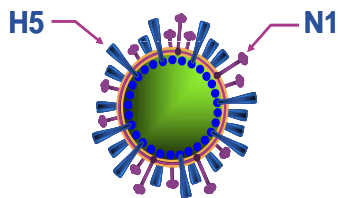
Data challenge on **EGEE**,
Auvergrid, **TWGrid**
 ~6 weeks on ~2000 computers

Target (**PDB**) :
 Neuraminidase (8 structures)

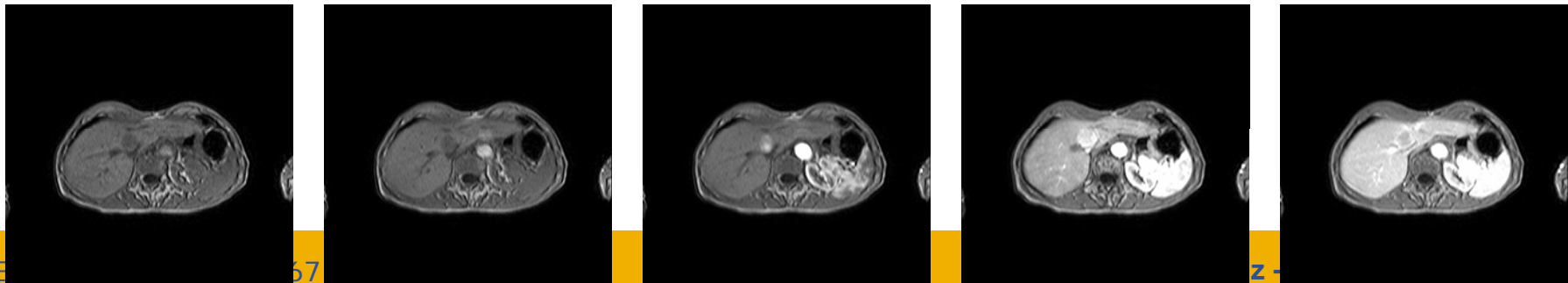
Hits sorting and refining

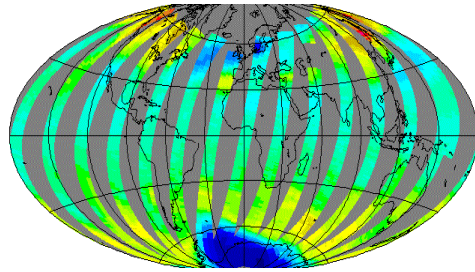


In vitro screening of 100 hits

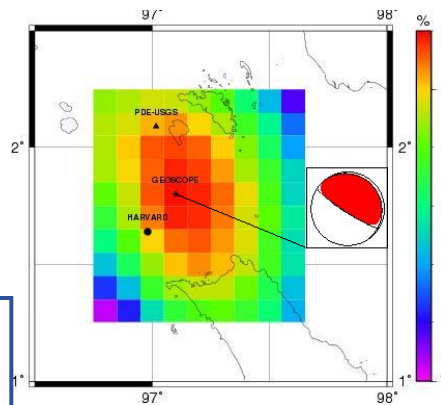


- **A lesion is detected in an MRI study of a patient**
 - **start with virtual biopsy**
 - The process requires obtaining a sequence of MRI volumetric images.
 - Different images are obtained in different breath-holds.
 - Before analyzing the variation of each voxel, images must be co-registered to minimize deformation due to different breath holds.
- **The total computational cost of a clinical trial of 20 patients is around 100 CPU days.**





ESA, UTV(IT),
KNMI(NL), IPSL(FR)-
Production and
validation of 7 years of
Ozone profiles from
GOME

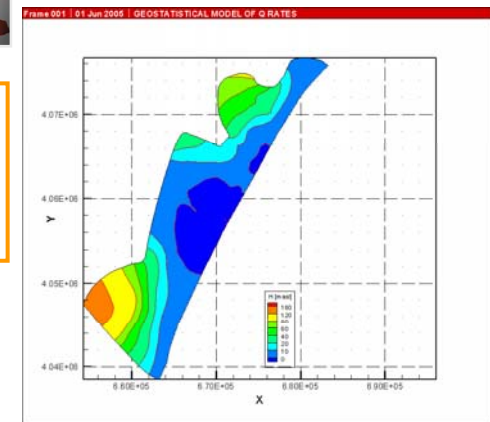


Rapid Earthquake
analysis
(mechanism and
epicenter)
50- 100CPUs
IPGP(FR)



Flood of a Danube river-
Cascade of models
(meteorology,hydraulic
,hydrodynamic....)
UISAV(SK)

Geocluster for
Academy and
industry CGG(FR)



Data mining
Meteorology &
Space Weather
(GCRAS, RU)

Modelling seawater
intrusion in costal
aquifer (SWIMED)
CRS4(IT),INAT(TU),
Univ.Neuchâtel(CH)

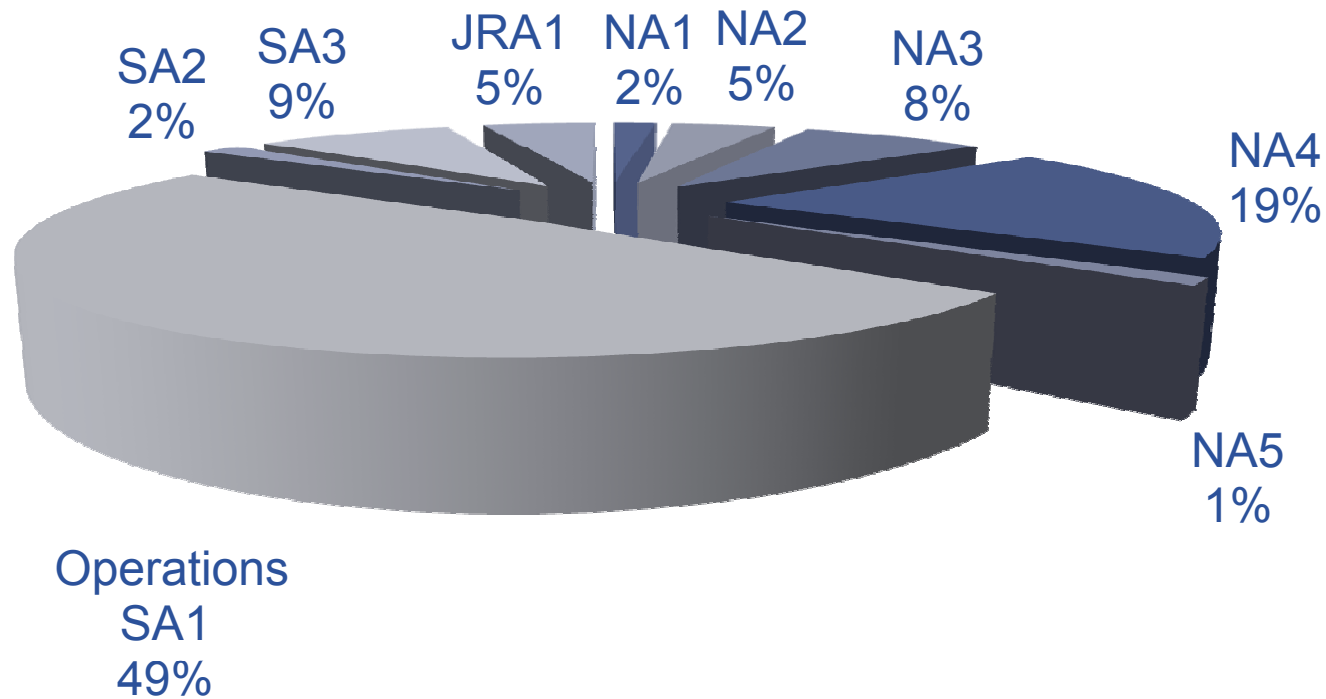
DKRZ(DE)- Data access
studies, climate impacts on
agriculture

Mars atmosphere CETP
(FR)

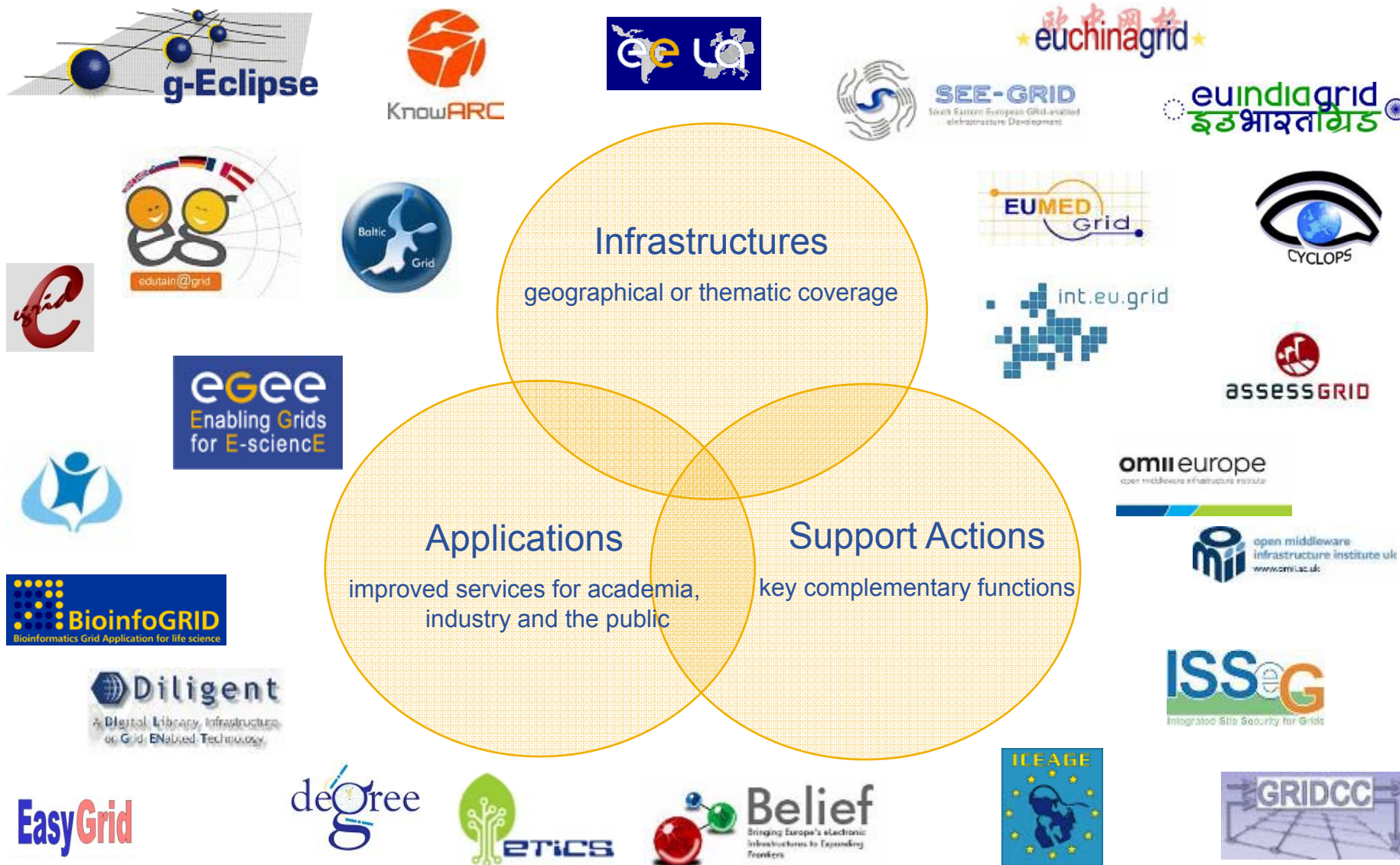
Specfem3D:
Seismic
application.
Benchmark for
MPI (2 to 2000
CPUs) (IPGP,FR)

Air Pollution
model- BAS(BG)

- **Total of 375 FTEs in EGEE-III**
 - 9010 person months (vs. 11165 PMs in EGEE-II; ~20% less)



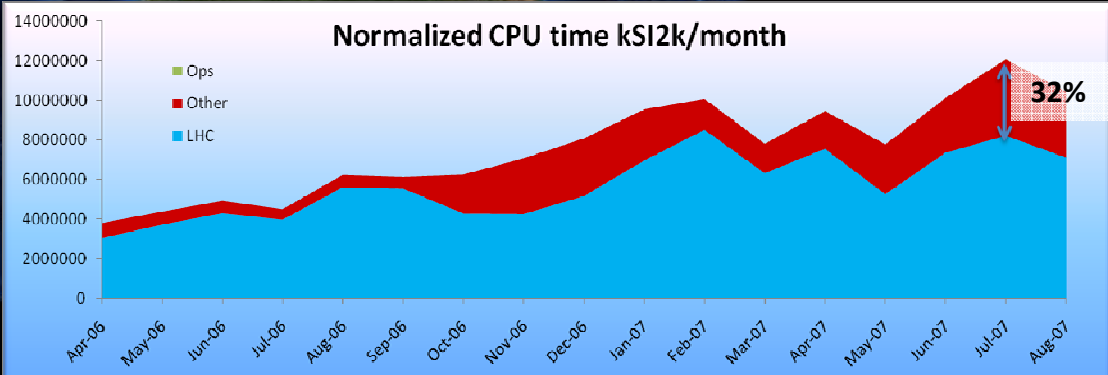
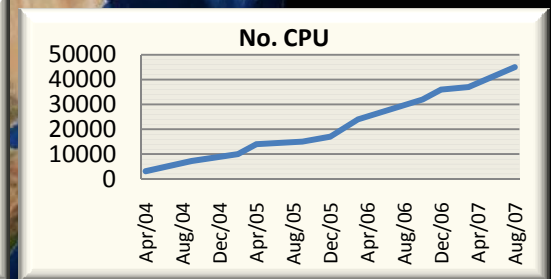
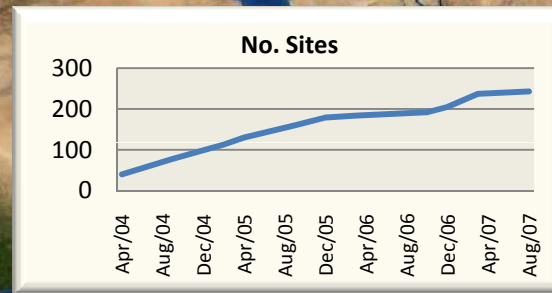
25 projects have registered as of September 2007: [web page](#)

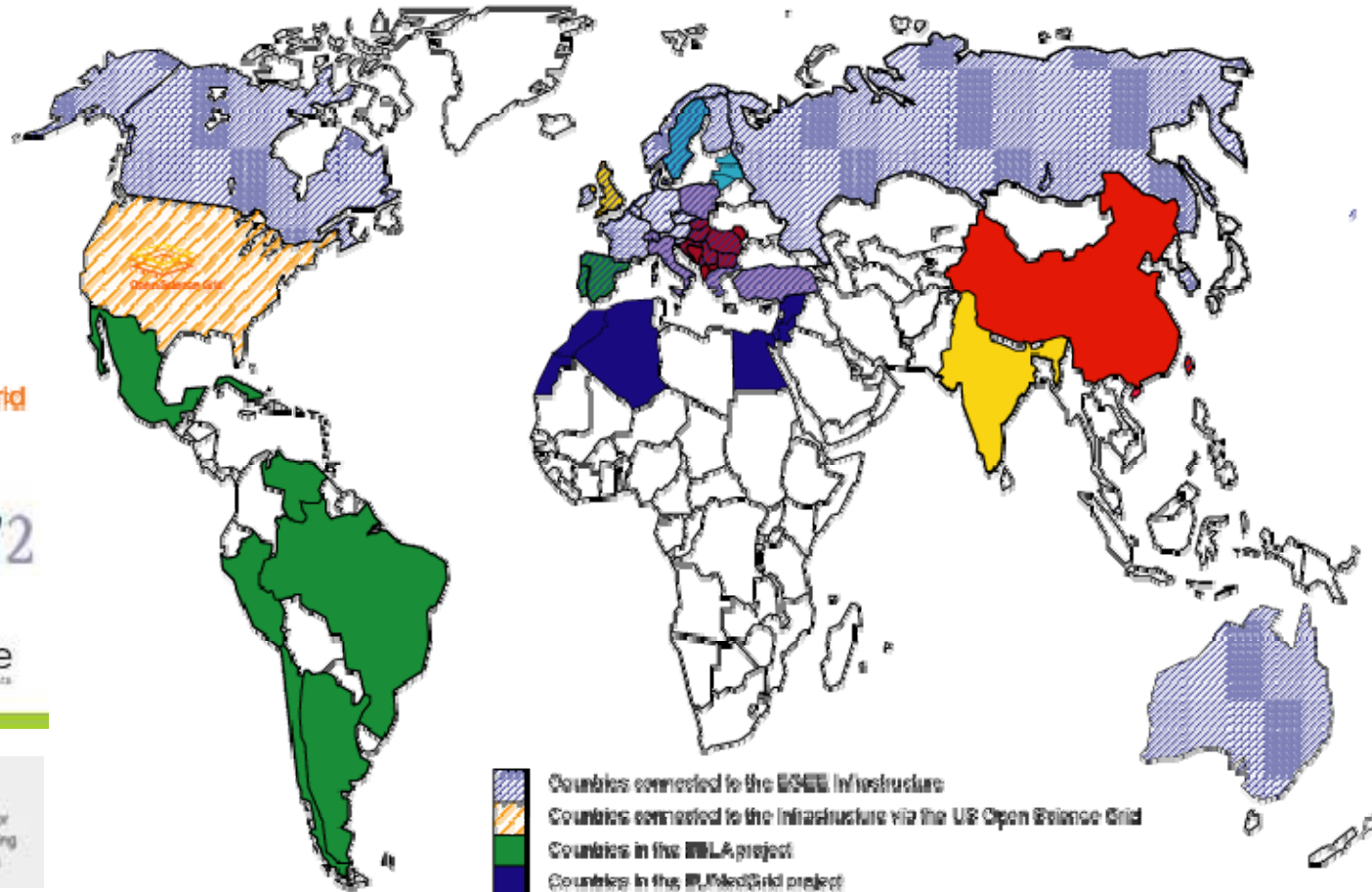


250 sites
48 countries
50,000 CPUs
13 PetaBytes
>5000 users
>200 VOs
>140,000 jobs/day

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences

...

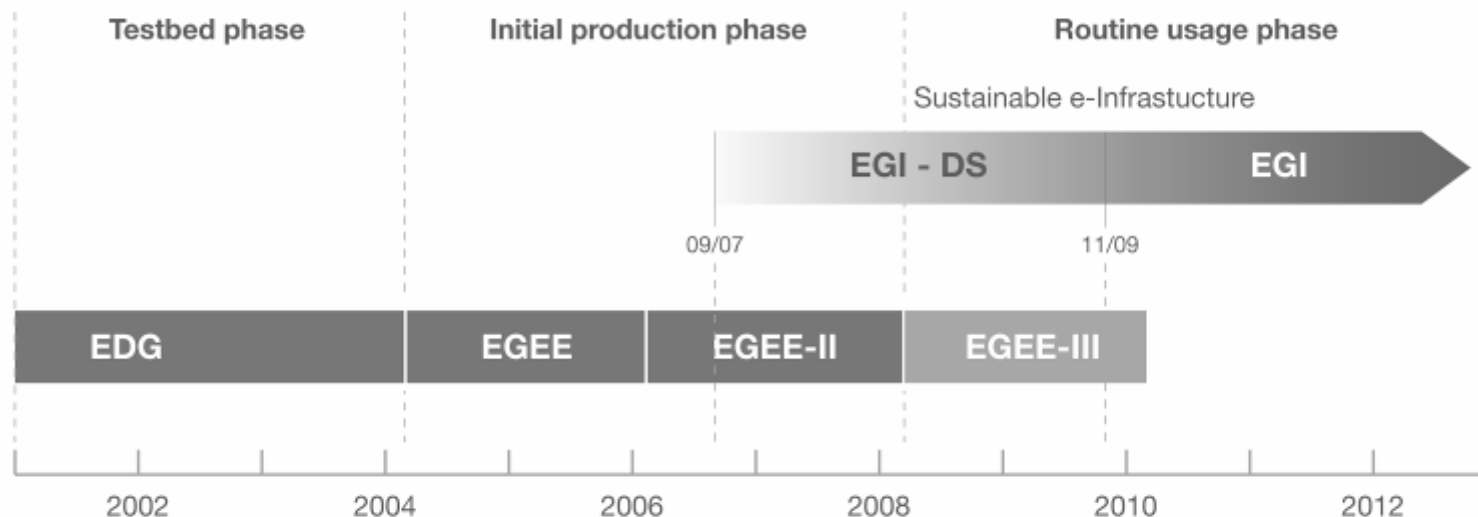




-  Countries connected to the EGEE infrastructure
-  Countries connected to the infrastructure via the US Open Science Grid
-  Countries in the EELA project
-  Countries in the EUMedGrid project
-  Countries in the BattleGrid project
-  Countries in the SIB-GRID project
-  Countries in the EUMindGrid project
-  Countries in the EUCChinaGrid project
-  Countries in several regional projects

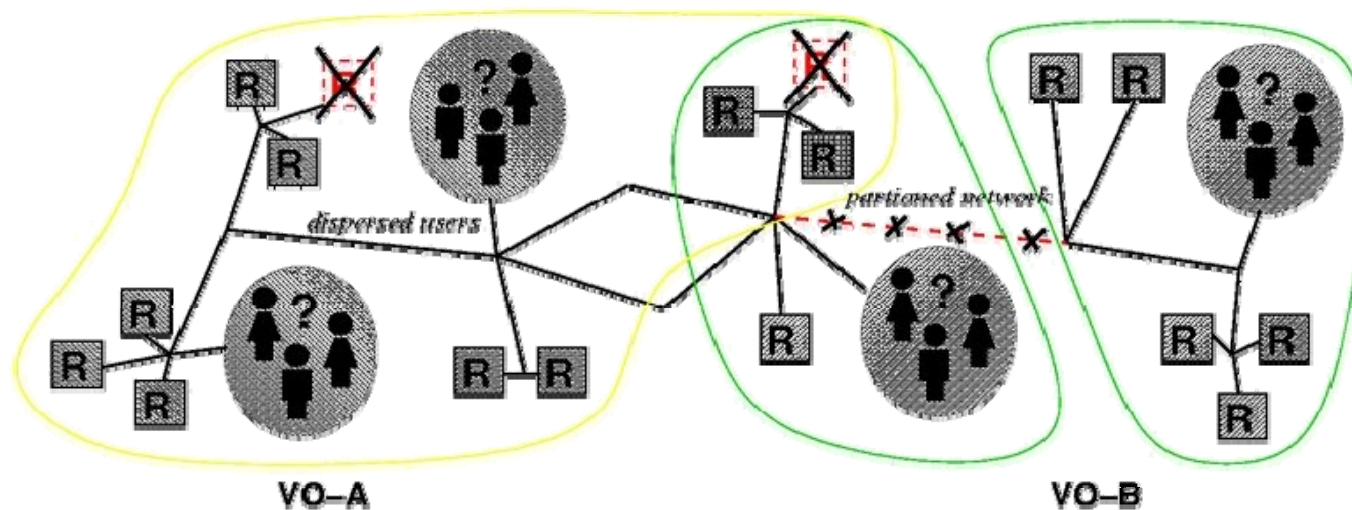


- Need to prepare permanent, common **Grid infrastructure**
- Ensure the long-term sustainability of the European e-Infrastructure independent of short project funding cycles
- Coordinate the integration and interaction between National Grid Infrastructures (NGIs)
- Operate the production Grid infrastructure on a European level for a wide range of scientific disciplines



What is a Virtual Organisation (EGEE take) ?

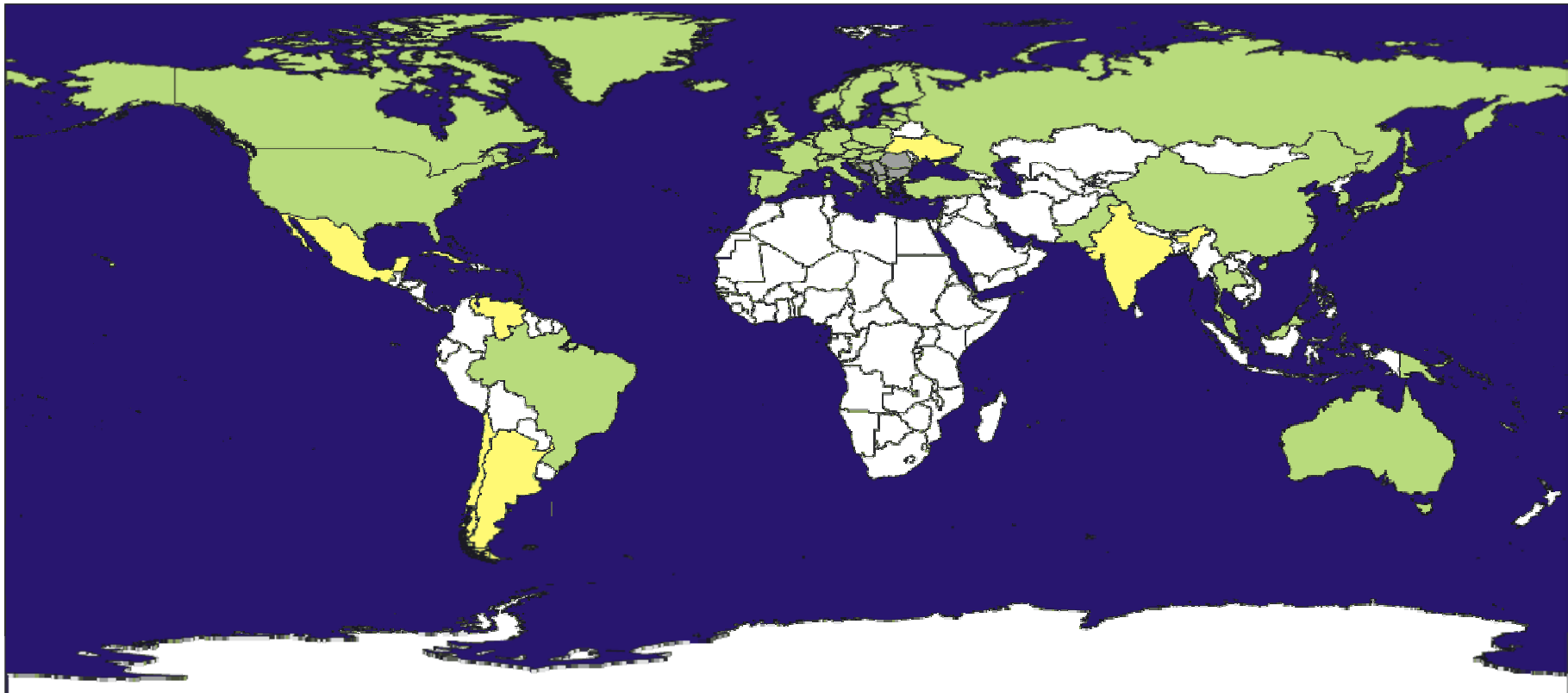
A set of individuals or organisations, **not under single hierarchical control**, (temporarily) joining forces to solve a particular problem at hand, bringing to the collaboration a subset of their resources, sharing those **at their discretion and each under their own conditions.**



graphic from: *Anatomy of the Grid*, Foster, Kesselman and Tuecke

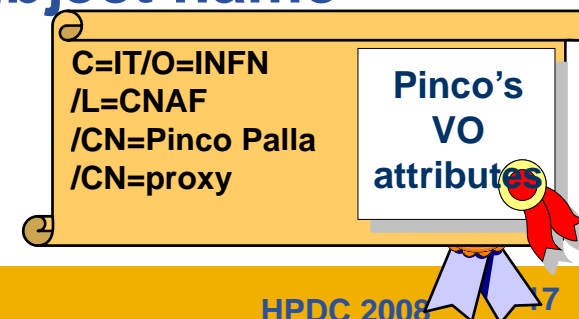
- **Within LCG/EGEE, VOs are essentially authorization domains:**
 - access rights to resources and datasets owned by a group of people
- **Authentication with X509 certificates**
- **Trust provided by IGTF**
- **Authorization with VOMS**
- **VO membership, group and role determines which resources (storage, computes) one has access to**
- **→ talk by Erwin Laure**

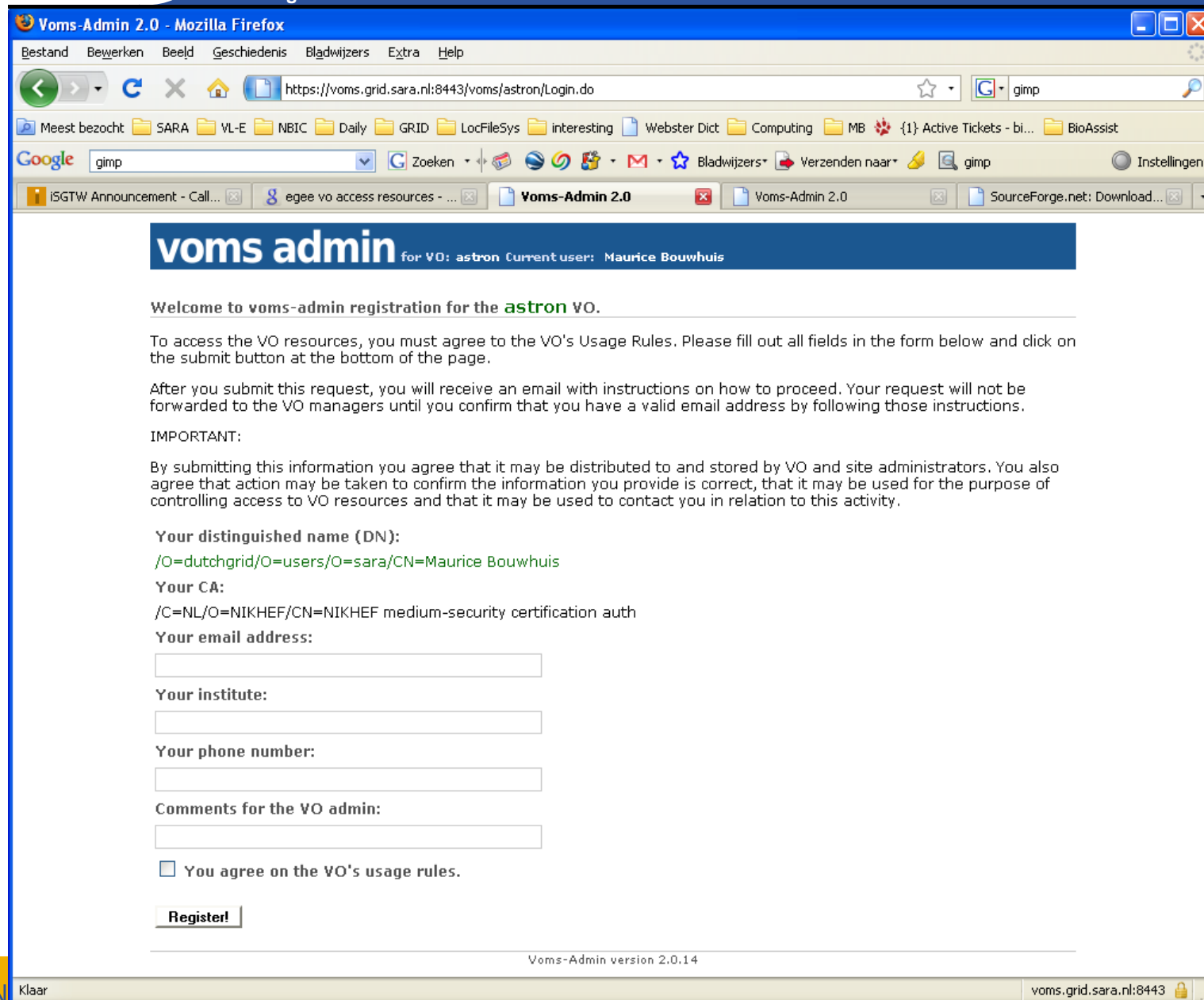
- All research grid infrastructures share the same base set of trusted third parties ('CAs')
- There is typically one in each country
- The credentials they issue are comparable in quality



- **Who becomes a member ?**
 - Commonality, they want something, they do not own themselves
 - Separation, they are not together physically/organizationally
 - Within EGEE adherence to similar community
- **Examples**
 - High Energy Physics people from Atlas storing on wLCG
 - Two theoretical chemists simulating on a regional grid

- If X509 is your passport then VO membership is the visa
- **Per-VO Authorisations (“visa”)**
 - granted to a person or service by a virtual organisation
 - based on the ‘passport’ name
 - acknowledged by the resource owners
 - providers can still ban individual users, and decide which privileges are granted to which VO attributes
- In your case, these ‘visa’ are called *VOMS credentials*
- It is a cryptographically protected statement by the VO
- which is bound (by the VO) to your subject name
- Roles and VO-groups are there





voms admin for VO: **astron** Current user: **Maurice Bouwhuis**

Welcome to voms-admin registration for the **astron** VO.

To access the VO resources, you must agree to the VO's Usage Rules. Please fill out all fields in the form below and click on the submit button at the bottom of the page.

After you submit this request, you will receive an email with instructions on how to proceed. Your request will not be forwarded to the VO managers until you confirm that you have a valid email address by following those instructions.

IMPORTANT:

By submitting this information you agree that it may be distributed to and stored by VO and site administrators. You also agree that action may be taken to confirm the information you provide is correct, that it may be used for the purpose of controlling access to VO resources and that it may be used to contact you in relation to this activity.

Your distinguished name (DN):
/O=dutchgrid/O=users/O=sara/CN=Maurice Bouwhuis

Your CA:
/C=NL/O=NIKHEF/CN=NIKHEF medium-security certification auth

Your email address:

Your institute:

Your phone number:

Comments for the VO admin:

You agree on the VO's usage rules.

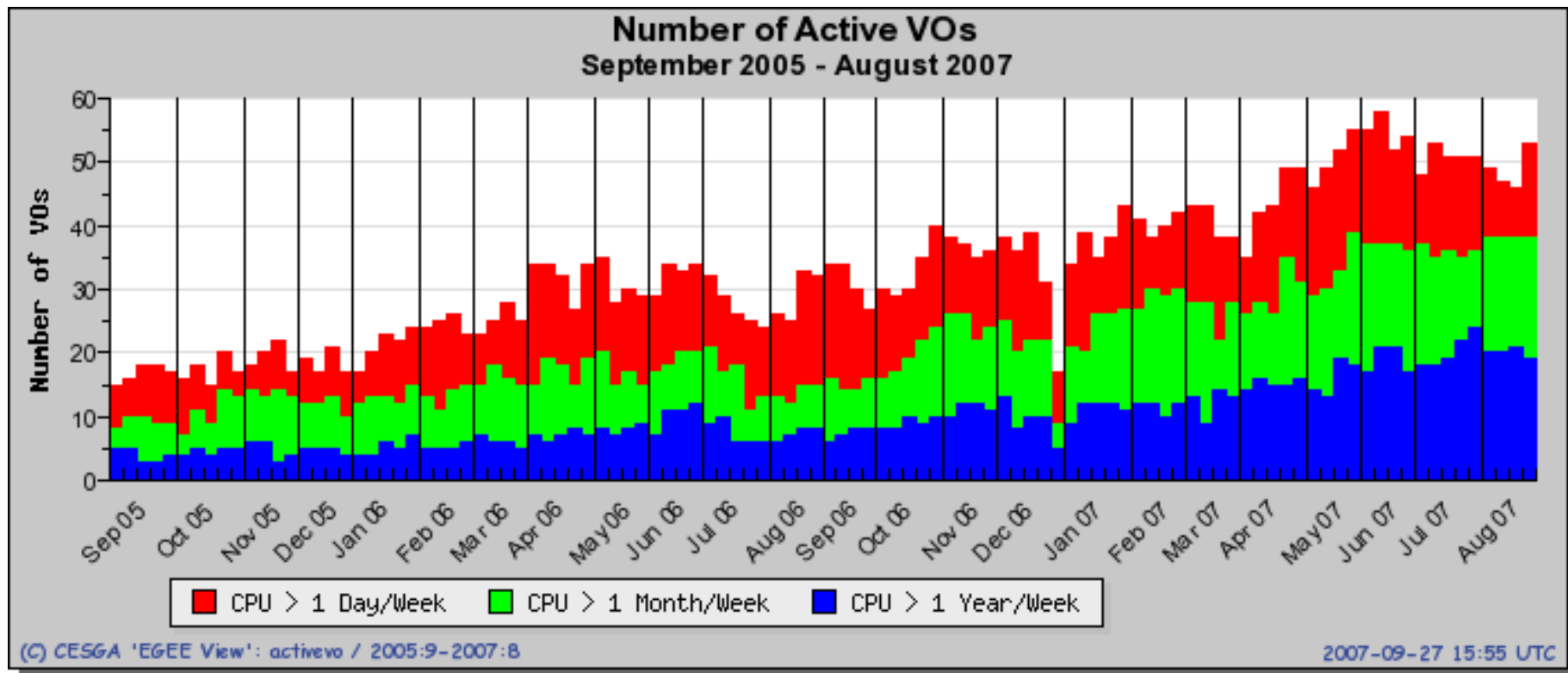
Register!

Voms-Admin version 2.0.14

- **Typical VOs that we manage**
 - State:
 - Registered VO, Adherence to EGEE policies
 - Selected/Supported VO, support from EGEE project
 - External VO, funded/supported by other project

 - Geographic Distribution
 - Local VO, at a single Resource Center
 - Regional VO, with a federation or country
 - Global VO, EGEE wide

- **Number of “active” VOs growing steadily!**
 - Turnover: Diff. VOs in last 6 / 12 / 24 months = 83 / 92 / 102
 - Total VOs: 104 registered, 258 visible



- **New Virtual Organisation (if no existing fits)**

- Setup a VOMS instance
- Be accepted by at least one Resource Center
- Access to Resource Broker
- Access to File Catalogue
- Register with EGEE portal
- EGEE compliant security policy (standard by EGEE) and acceptable use Policy

Usually one of the Regional Centers will provide these services

Not a really dynamic process, but fits the e-Science requirements

- **More Resources**

- For EGEE-wide VOs negotiate through Regional Operations Centers
- Regional VOs can usually join regional grid infrastructures
- **Always** needs an action and decision form each resource provider

- **Sites in EGEE usually support many VOs**
- **Ties between each VO and sites are loose**
 - Sites must be more generic in their setup
 - VOs must program with more discipline
- **Each VO has role “Software Admin” and each site provides software area**
- **In LCG exist VO-boxes for site related VO specific services**

- **EGEE provides a dependable production quality Grid infrastructure to a wide variety of scientific disciplines.**
- **Grids are increasingly becoming an essential part of the scientific computing infrastructure – sustainability needs to be ensured**
- **VO setup is static**
- **VO membership (incl. group and role determine access)**