



Enabling Grids for E-scienceE

Use of EGEE Infrastructure Evolution of NA4 Activity

V. Floros (GRNET), C. Loomis (CNRS/LAL)

*EGEE'07 Conference
Budapest, Hungary
1-5 October 2007*

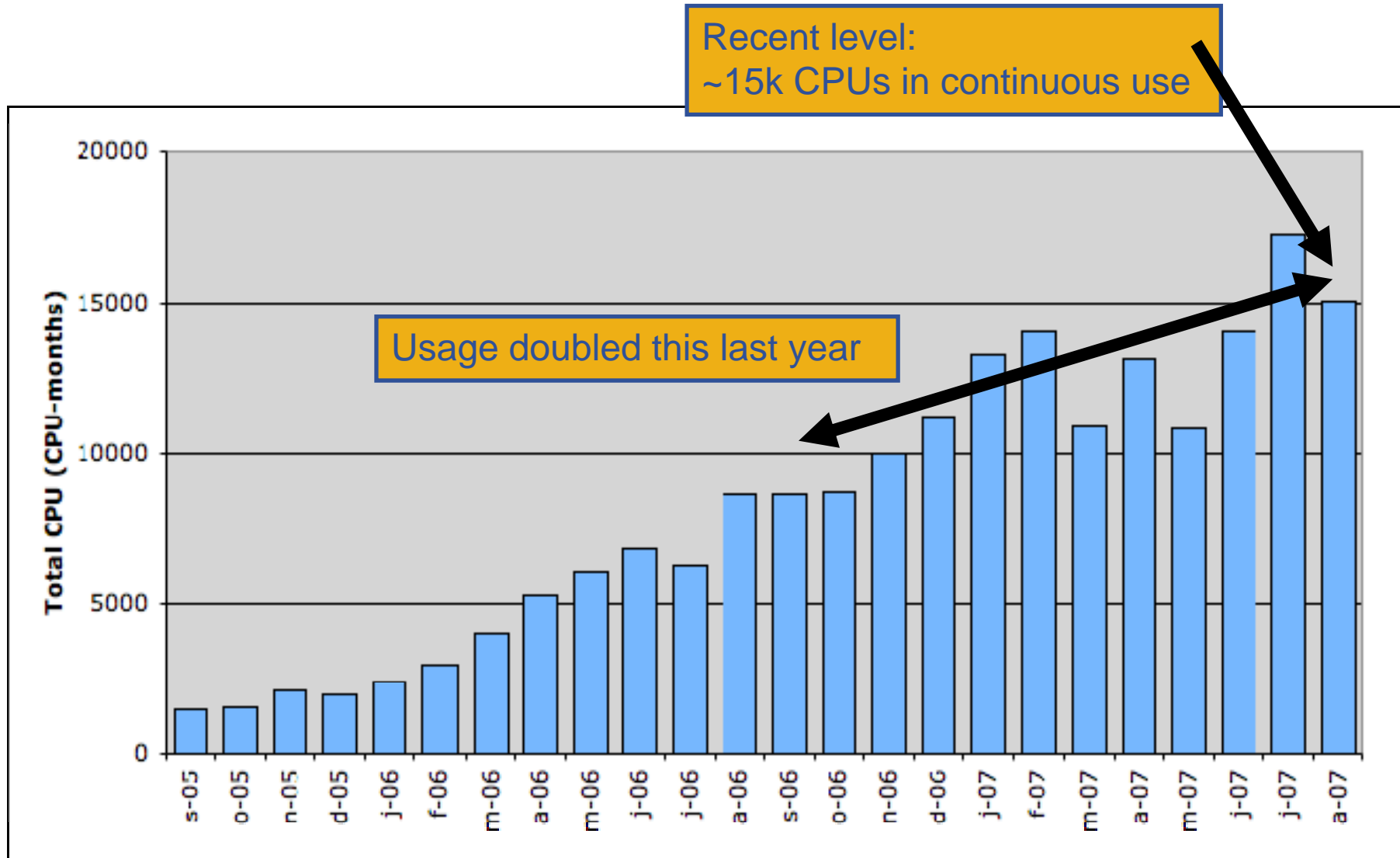
www.eu-egee.org



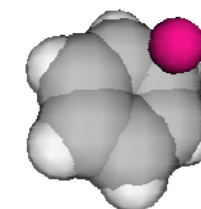
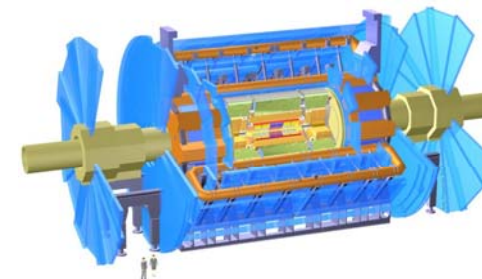
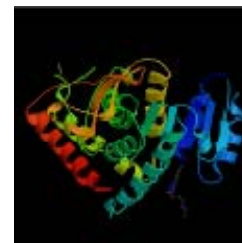
- **Introduction**
 - Goals of NA4.
 - Why use the grid?
- **Use of the EGEE infrastructure**
 - Statistics
 - Highlighted applications
- **Foreseen evolution**
 - Sustainable support model
 - Services (to be) provided
- **Summary**

Statistical information obtained from:
EGEE Accounting Portal
Information System
CIC Portal

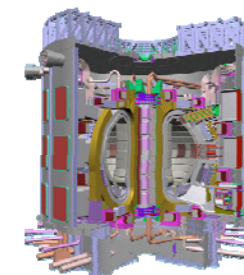
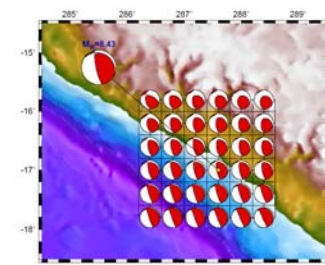
- **Goal: More, happy users!**
- **Expand use of EGEE infrastructure:**
 - Virtual Organization: More experiments, collaborations.
 - Users: More scientists from each experiment.
 - Applications: More, more variety of applications.
- **Ensure current users are satisfied.**
 - Support through all stages of use.
 - Provide opportunities for feedback.
- **Collaboration**
 - Platform to bring different people with different skills together.
 - Mechanism to analyze, publish, and combine previous results.



- Disciplines: 10
- Sub-disciplines: 36
- See growth and diversification of applications.
- Reported apps. only \Rightarrow *underestimate!*

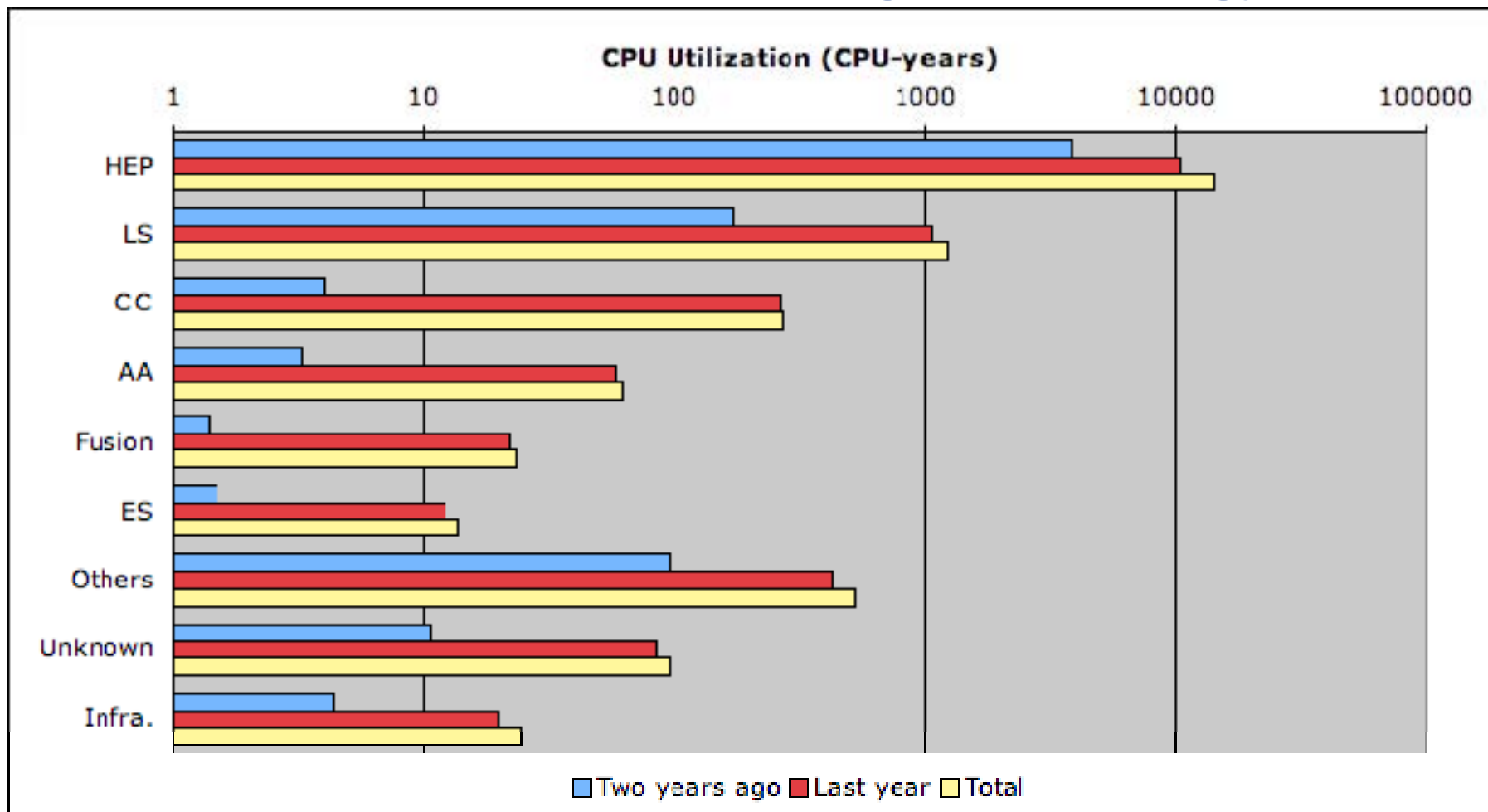


	6/2006	2/2007
Astronomy & Astrophysics	2	8
Computational Chemistry	6	27
Earth Science	16	16
Fusion	2	3
High-Energy Physics	9	11
Life Sciences	23	39
Others	4	14
Total	62	118



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

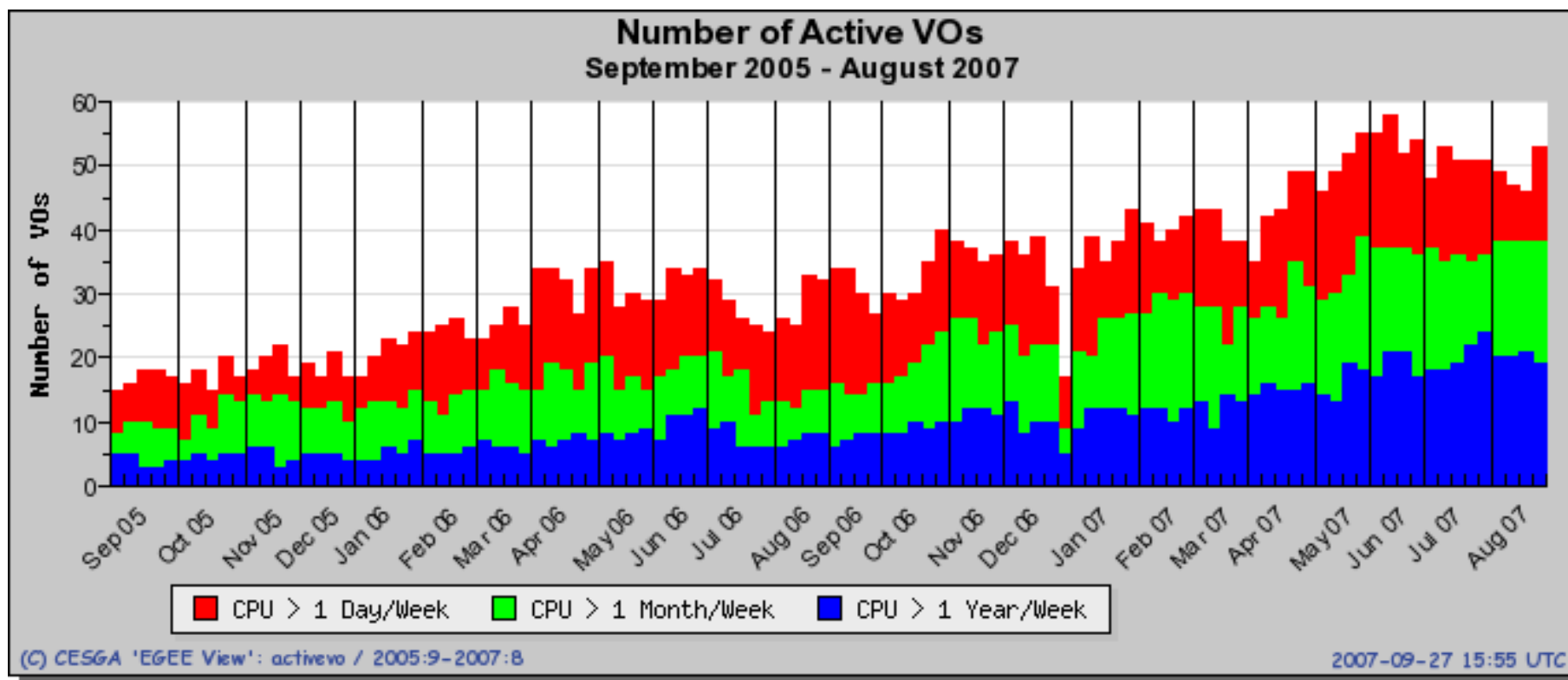
- Wide (natural) differences in total CPU utilization.
- Evidence of broad adoption of grid technology.



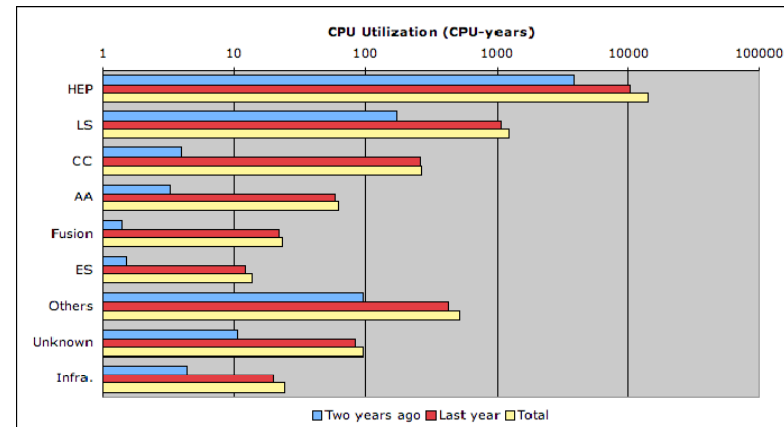
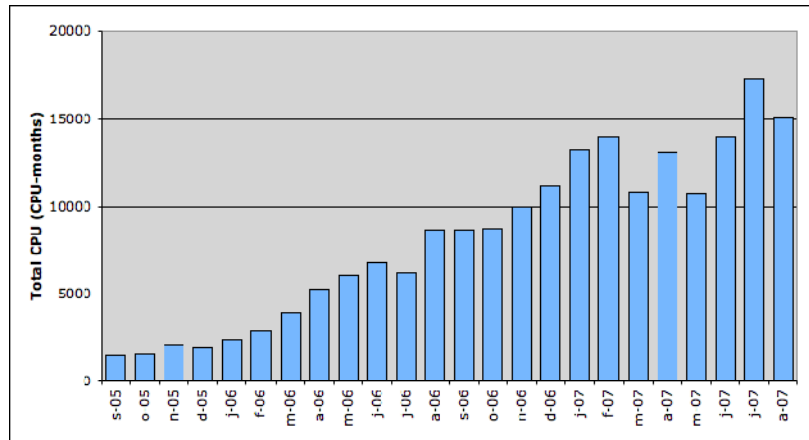
- Utilization depends on having available resources.
- See good coverage of scientific disciplines for computing and storage resources.
 - Sites often have more than one CE or SE defined.
 - Number *not* size of resources!
- Thanks!

	# CEs	# SEs
HEP	292	299
LS	113	123
CC	25	41
AA	57	83
Fusion	19	21
ES	42	65
Others	143	149
Unknown	288	327
Infra.	282	306
Total	366	334

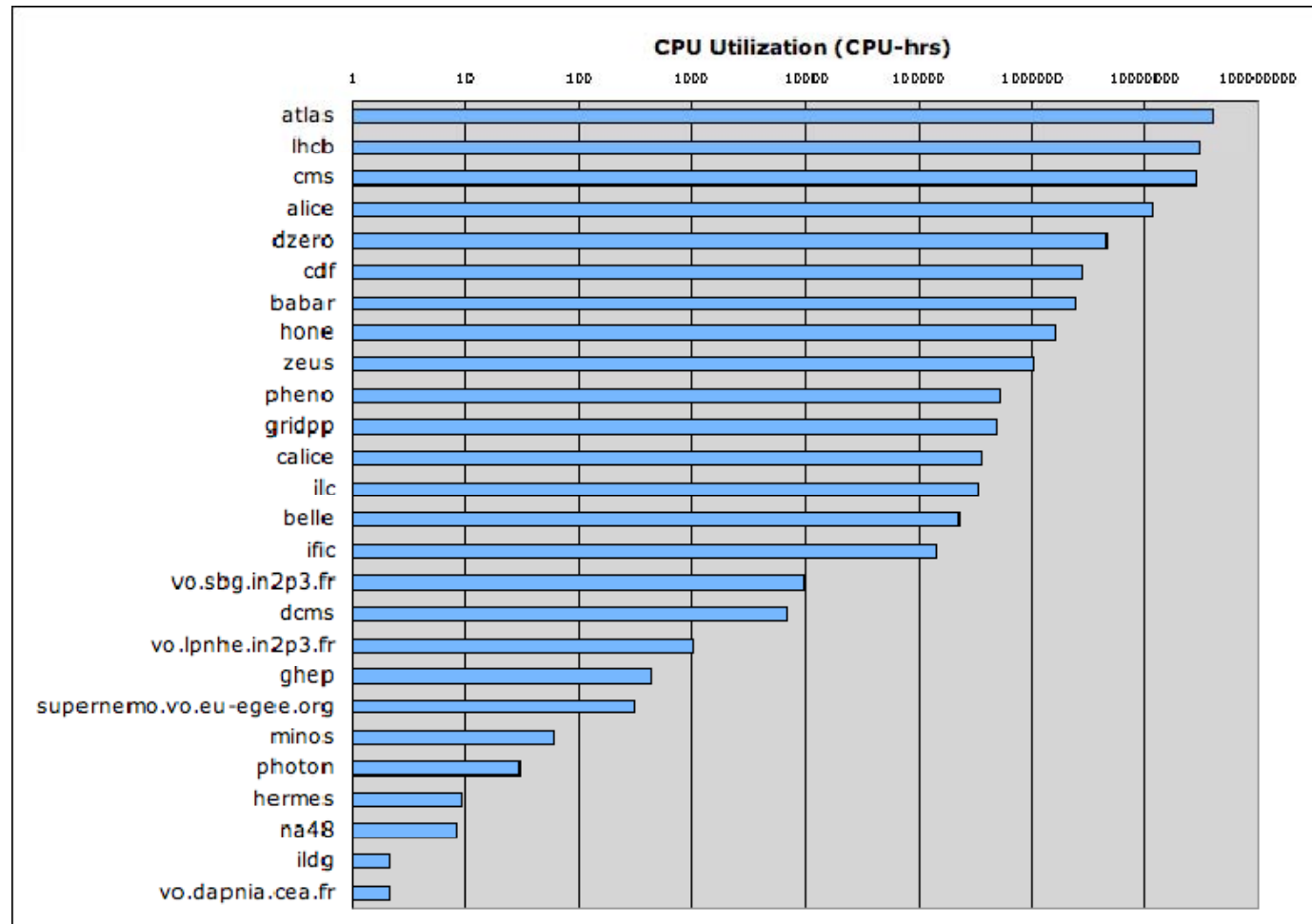
- **Number of “active” VOs growing steadily!**
 - Turnover: Diff. VOs in last 6 / 12 / 24 months = 83 / 92 / 102
 - Total VOs: 104 registered, 258 visible



- Large, growing overall utilization
- Long-term, habitual use of infrastructure.
- Broad adoption many diverse communities



- **Coordinator: Massimo Lamanna**
- **Use by major (& future) HEP laboratories**
 - LHC
 - Tevatron
 - Hera
 - SLAC
 - ILC
 - ...



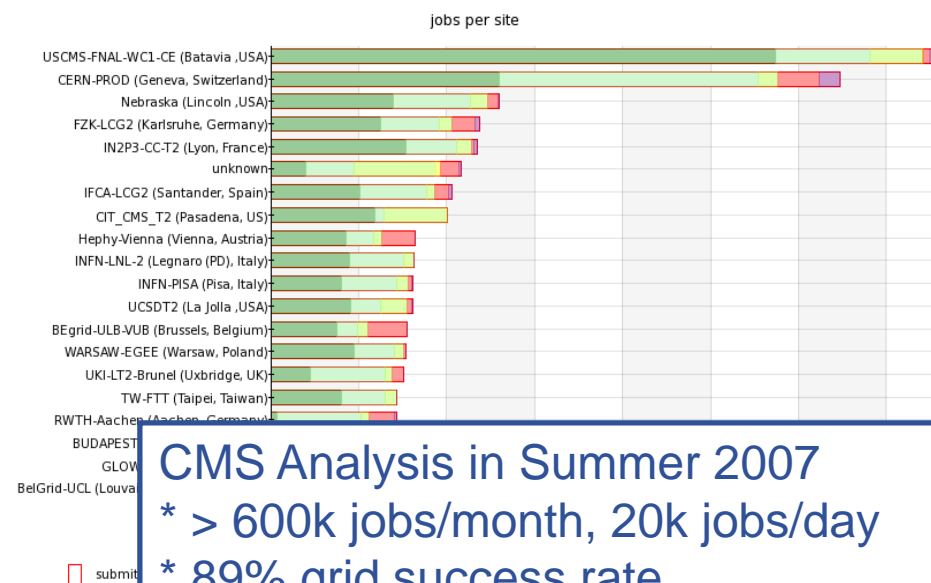
- **Dashboard: serving 4 of 4 LHC experiments**

Interest from other communities:

- * Pilot for Life Sciences (vlemed VO)
- * Interest from Diligent

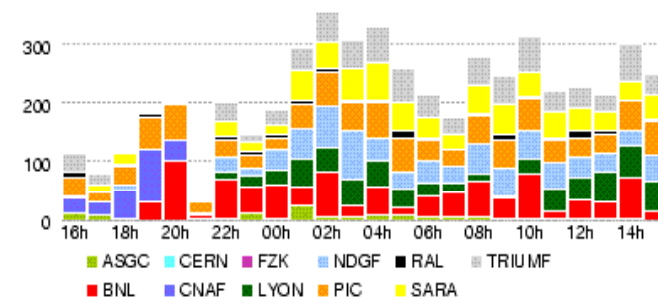


ATLAS collected and distributed first detector data (cosmic rays) to Tier2s



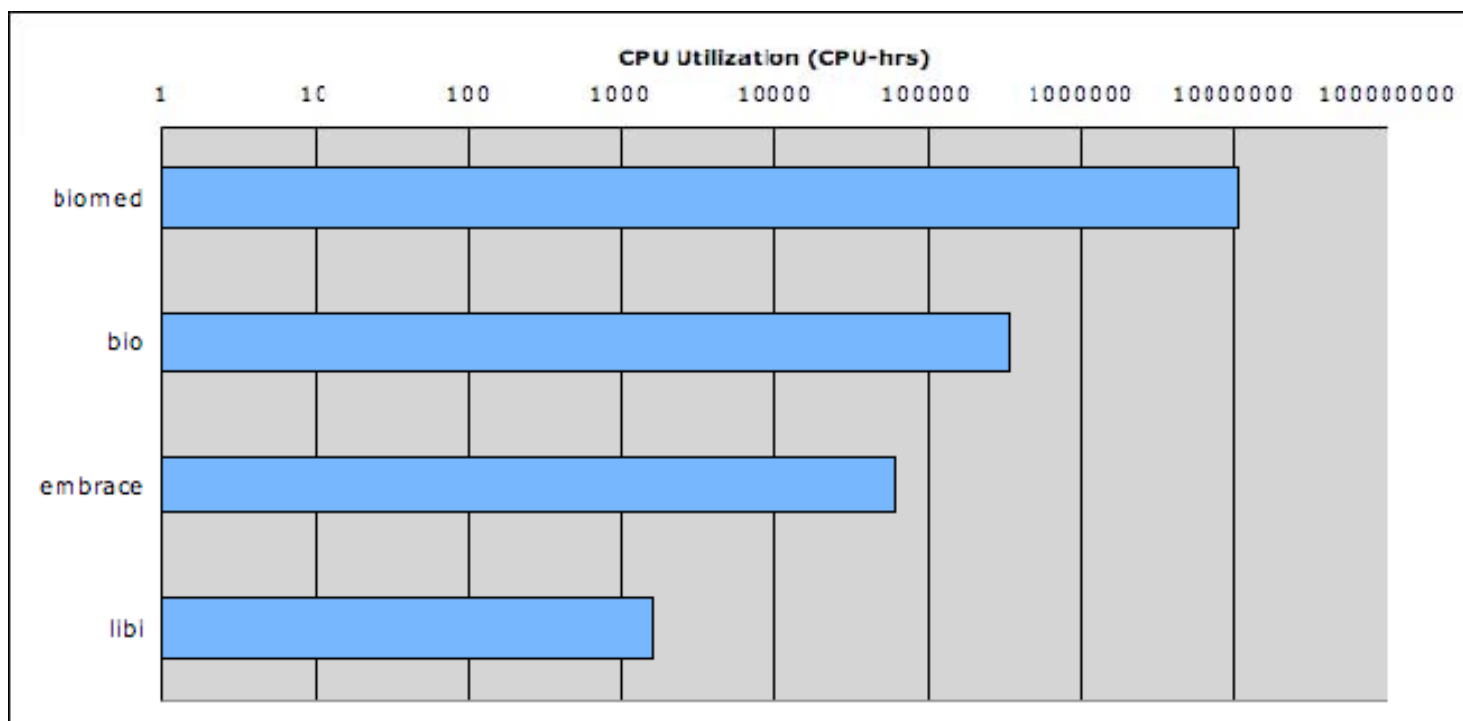
CMS Analysis in Summer 2007

- * > 600k jobs/month, 20k jobs/day
- * 89% grid success rate
- * substantial Tier2 contribution

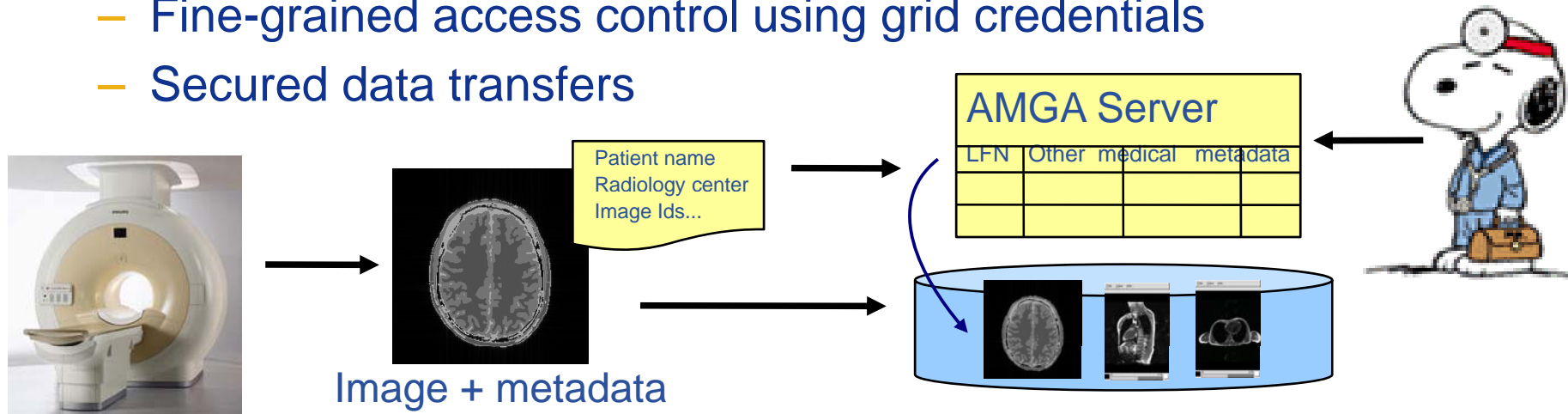


Contribution to monitor users' jobs:
Grid reliability + more information to final users

- **Coordinators:**
 - Drug Discover: Vincent Breton
 - Medical Imaging: Johan Montagnat
 - Bioinformatics: Christophe Blanchet

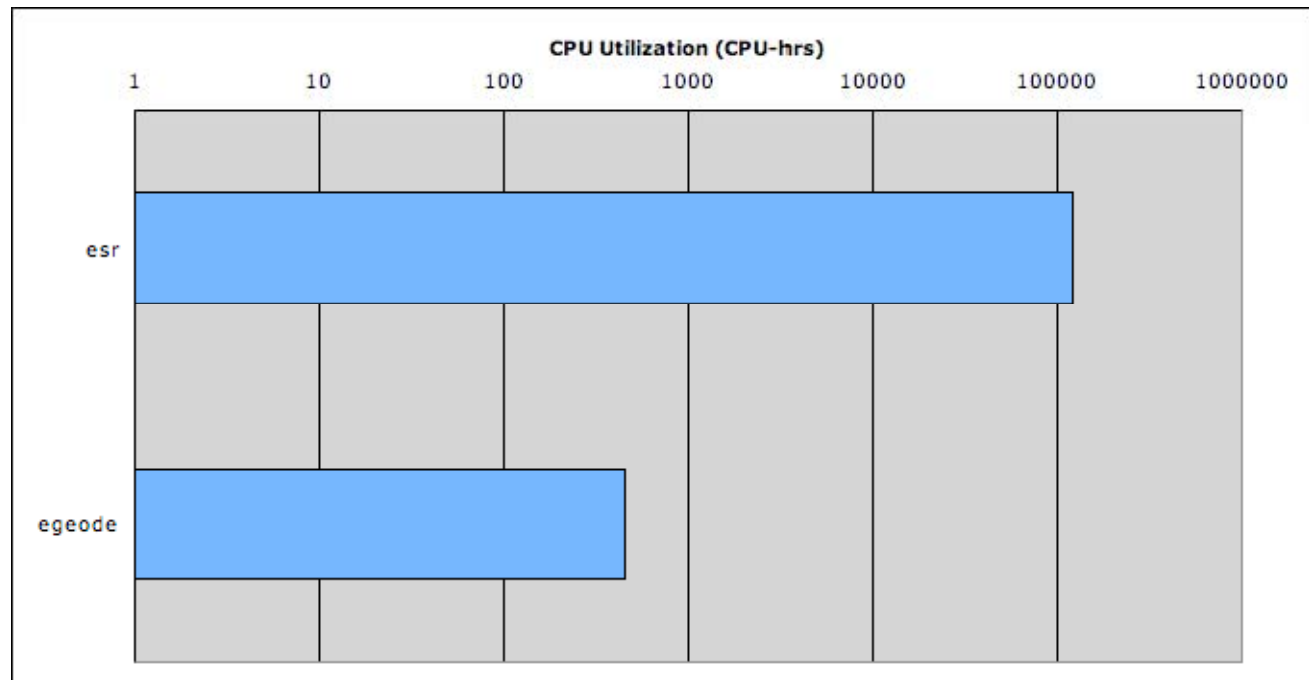


- **AMGA server used for medical metadata management**
 - Fine-grained access control using grid credentials
 - Secured data transfers



- **Medical data mgt. environments exploiting AMGA**
 - gLite Medical Data Manager (CNRS, I3S lab)
 - Medical image management web portal (CNRS, LPC)
 - Alzheimer's patient data analysis env. (Biolab, U. Genova)
 - Health-e-child Data Management System (Health-e-child project)
- **Future evolution: Distributed metadata repositories**

- **Coordinator: Monique Petitdidier**
- **Extremely varied range of applications in this sector.**



Sharing Algorithms

GEOCLUSTER

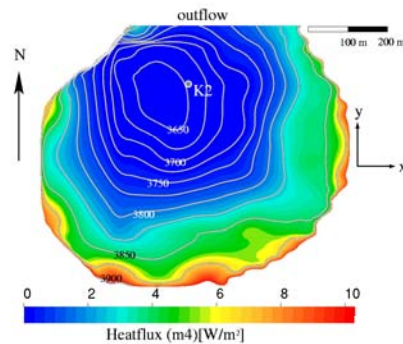
CGG-Veritas



Partners:
VO - EGEODE

ELMER

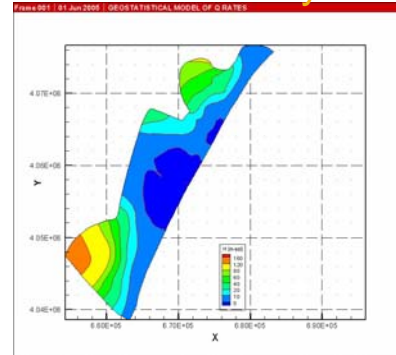
CSC - Finland



Partners:
VO - ESR

CODESA-3D

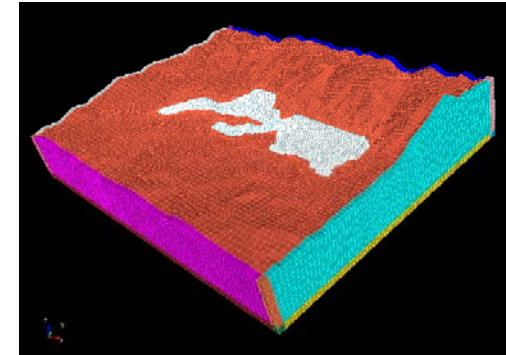
CRS4 - Italy



Partners:
EUMEDGrid

3DSEM_UNSTRUCT

IPGP- France



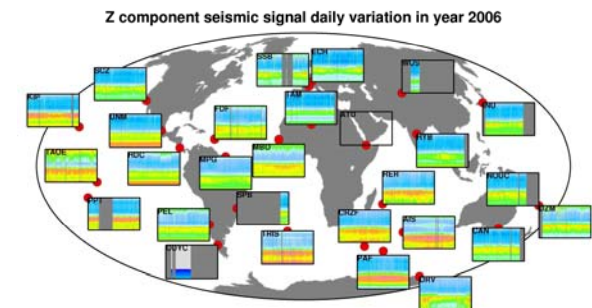
Partners:
EELA

Exploring Large Set of Data

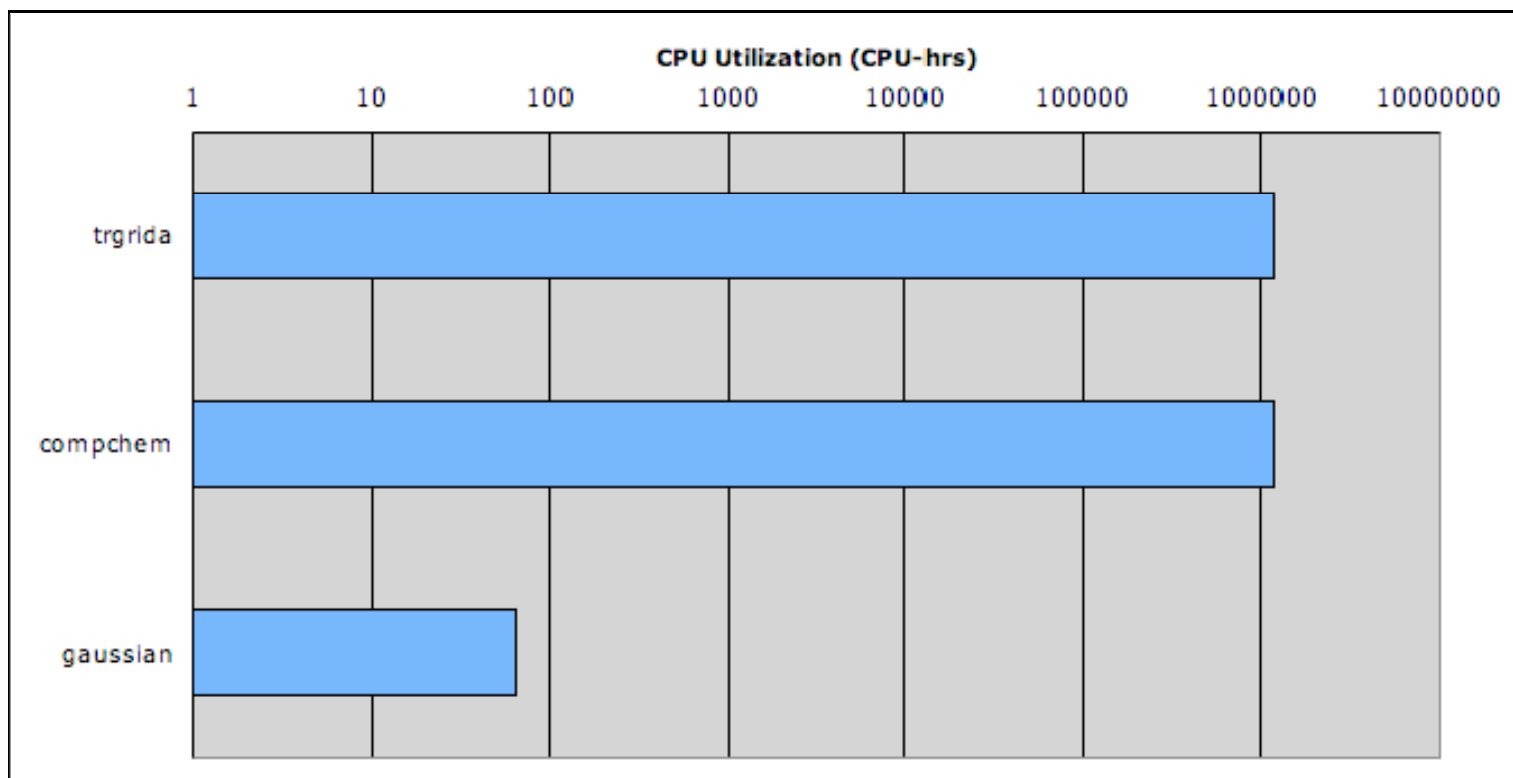
Geoscope: (<http://geoscope.ipgp.jussieu.fr>) IPGP-France

- 28 seismological stations and data center
- 25 years of data
- Processing of the whole data set on EGEE

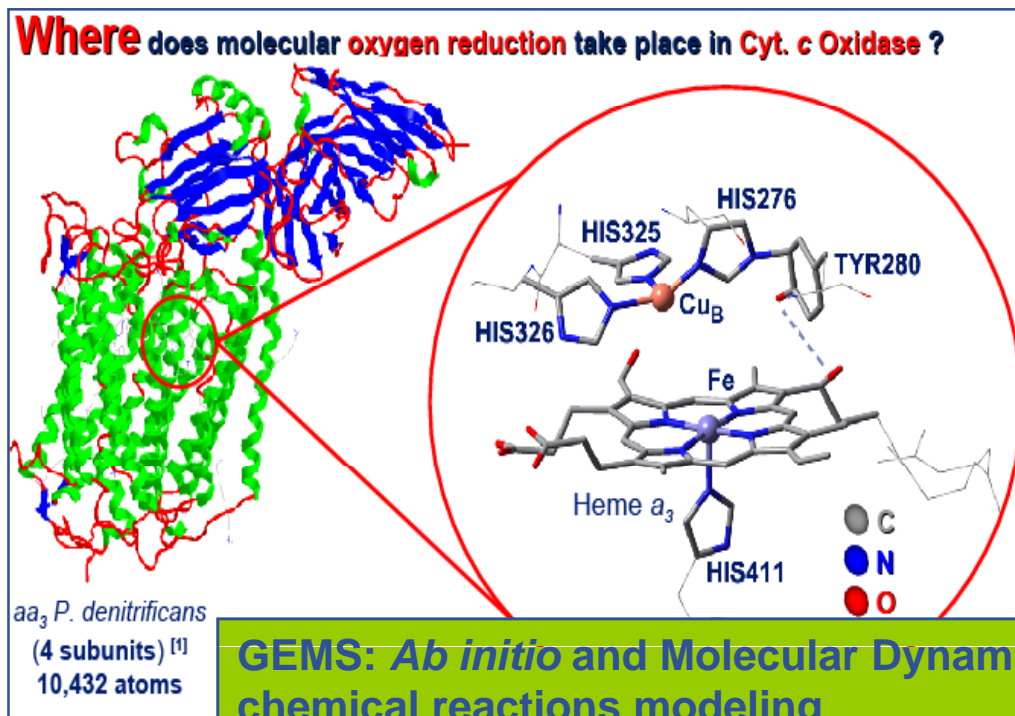
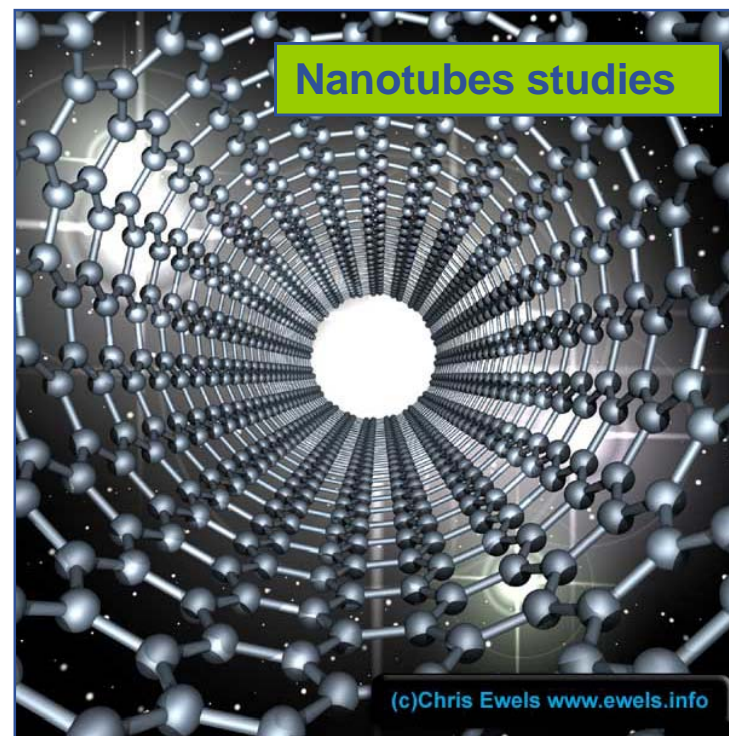
Impact on seismological data center design



- **Coordinator: Mariusz Sterzel**



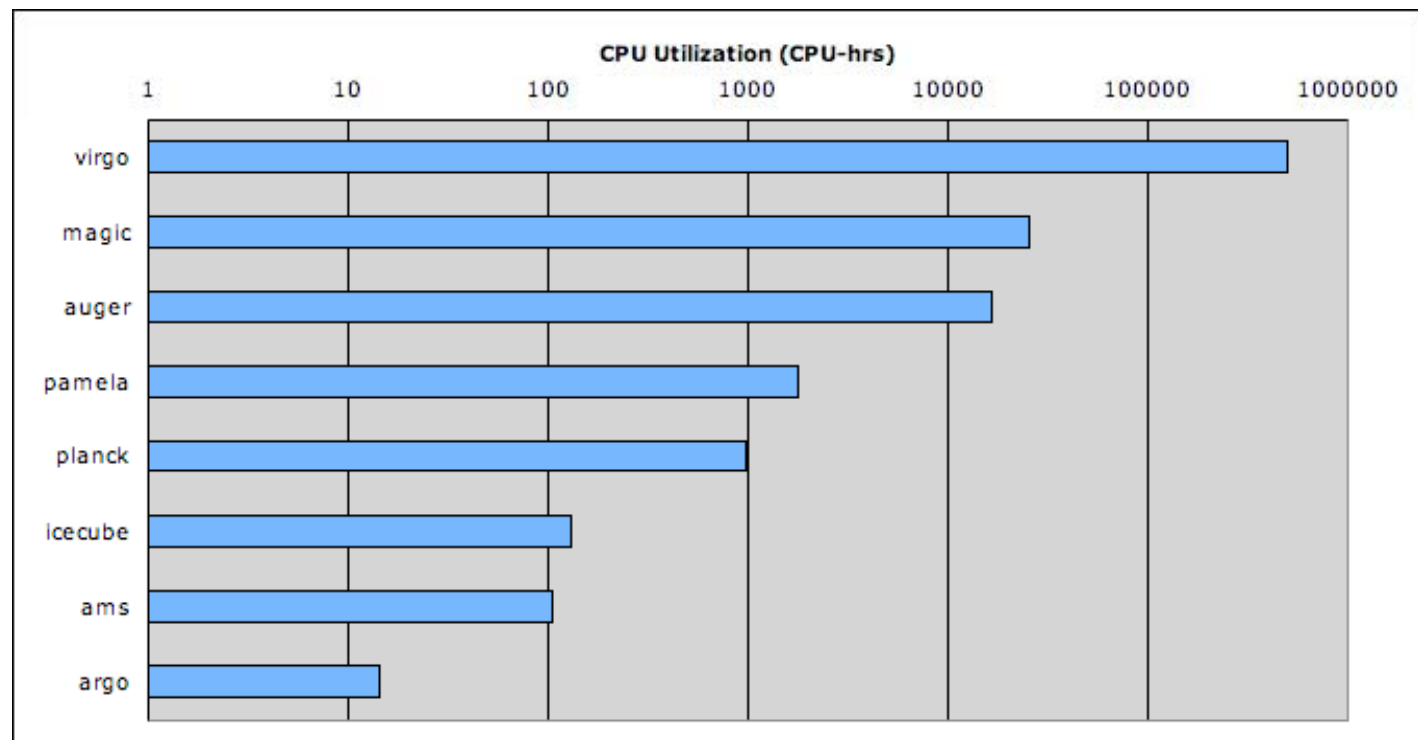
- **Commercial software availability**
 - Gaussian, Turbomole
- **Parallel (MPI) execution**
 - DL_POLY, NAMD, Turbomole



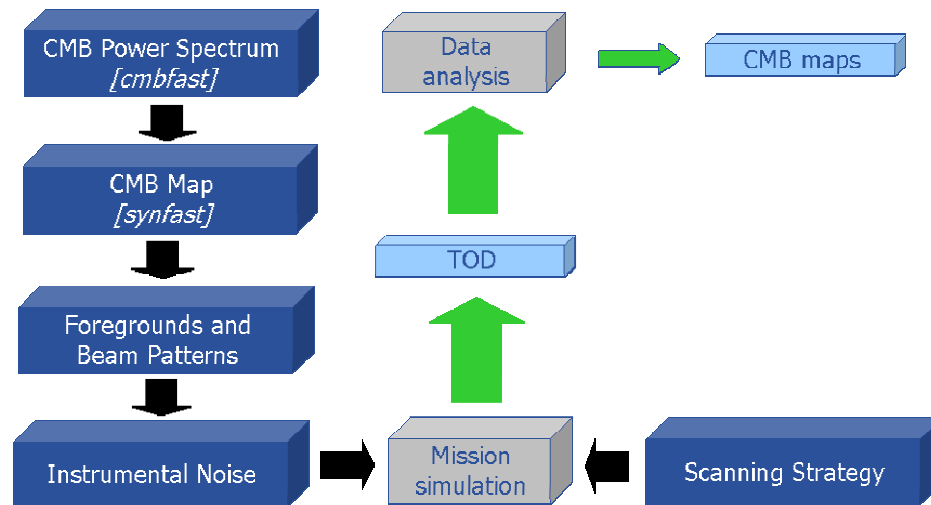
Other interested disciplines:

- * Biology
- * Pharmacology
- * Solid state physics

- **Coordinator: Claudio Vuerli**



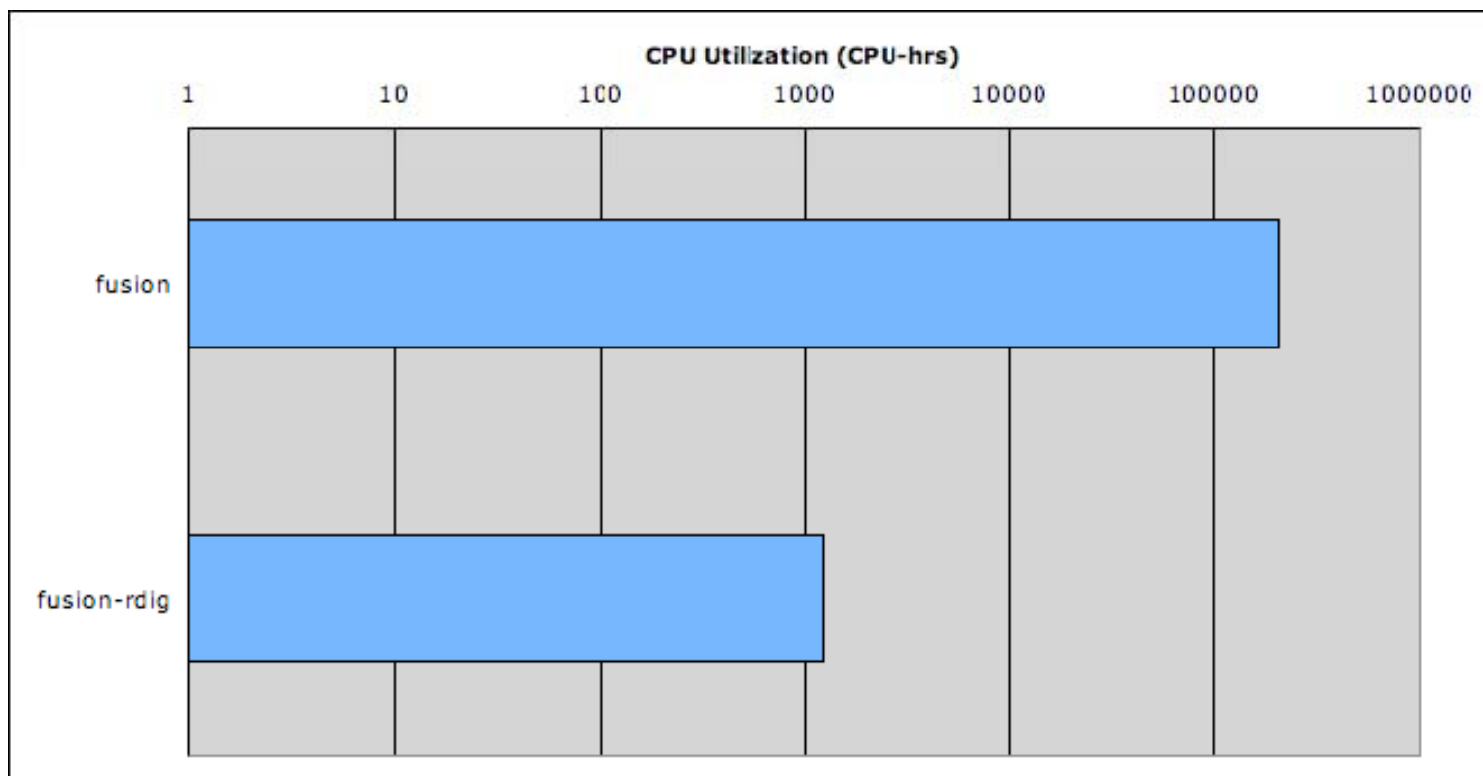
- Typical application that benefits by distributed computing techniques
- International collaboration with a high level of interactions.
- Usage of local clusters, EGEE infrastructure and DEISA facilities.



Requirements/Wishes

- Moving huge amount of data over the grid.
- On the fly deployment of code on the grid.
- Saving intermediate data directly on SEs.
- Deployment of dedicated libraries.
- Deployment of visualization tools.
- Make EGEE G-DSE compliant.
- Integration in grid portals (Genius, EnginFrame, etc.).
- More effort to train people.

- **Coordinator: Francisco Castejon**



- **Gaussian**

- <http://www.gaussian.com/>
- Predicts the energies, vibrational freq., ... of molecular systems.
- VO-based licensing model, actually in use in gaussian VO.

- **MathWorks**

- <http://www.mathworks.com/>
- Integrate MATLAB & Distributed Computing Engine with EGEE.
- Both client and server are licensed in this model.

- **Interactive Supercomputing**

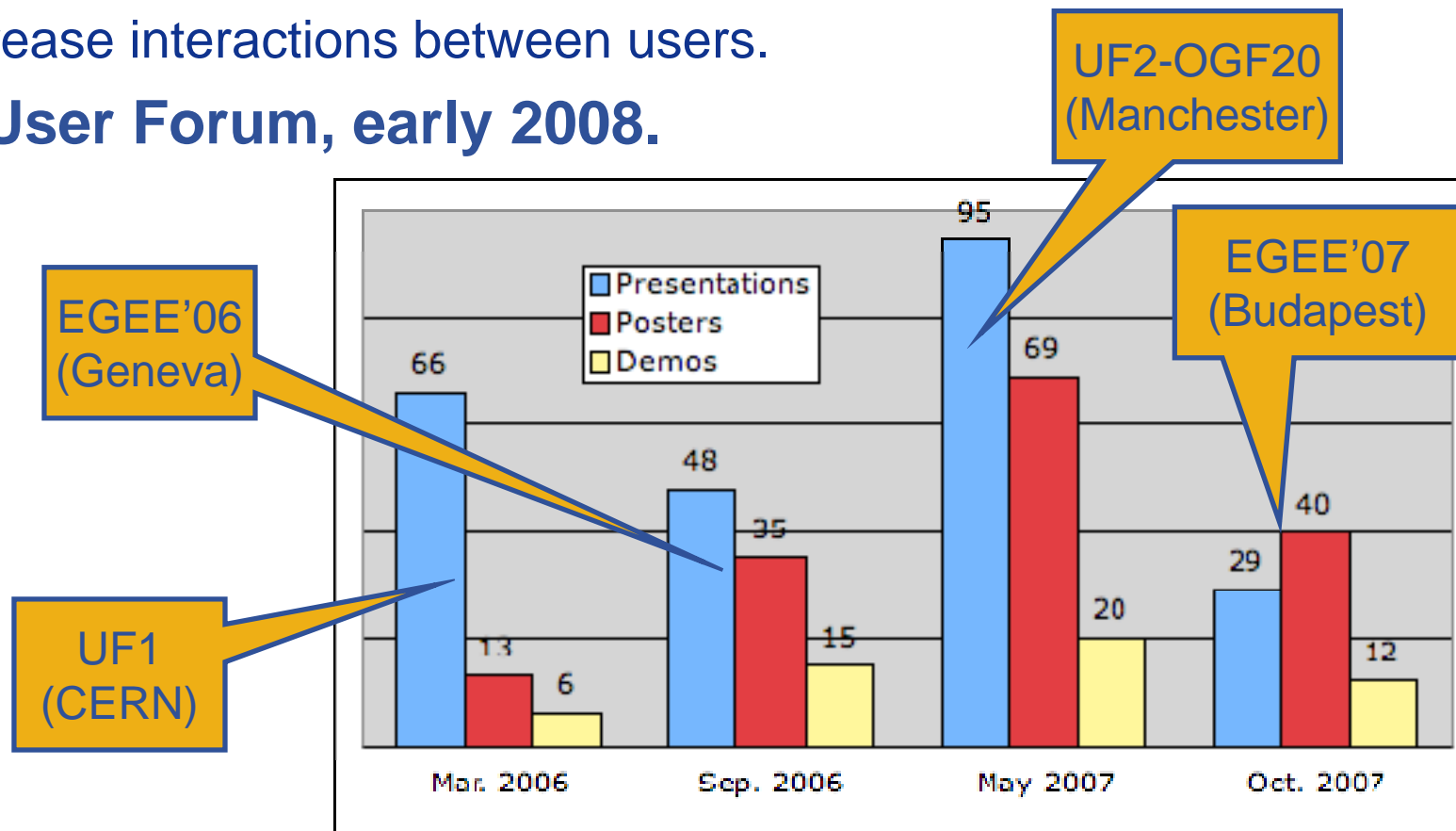
- <http://www.interactivesupercomputing.com/>
- Similar to DCE; used from multiple clients (MATLAB, Python, R)
- Server licensed, some clients licensed

- **Goal of evolution?**
 - Adopt a support (and service) model consistent with a sustainable grid infrastructure.
- **How?**
 - Provide critical user/community support services.
 - Interact with and take advantage of regional efforts.
 - Provide coordination between actors.
- **What?**
 - “Classic” Support
 - Community Building
 - Advanced functionality
 - Management

- **Ensure daily interactions with grid are effective.**
- **Direct User Support**
 - Continue to use GGUS system.
 - New team within NA4: handling tickets, resp. for documentation.
- **VO Support**
 - Continues through the VO Managers' Group (F. Schaer)
 - Registration, Treatment of VO issues
- **Application Porting Support**
 - GILDA: t-Infrastructure moves to NA3 activity
 - SZTAKI: New team (already started) provide additional support

- **Build strong, self-reliant user communities.**
- **Discipline-specific meetings**
 - Techniques to aid each discipline (common data, tools, etc.)
 - Dissemination within that discipline
- **Topical Meetings**
 - Discuss common problems or needs
 - Highlight tools/techniques to address those needs

- **Conferences = Knowledge Transfer**
 - Present results from using grid technology.
 - Discuss encountered problems and solutions.
 - Increase interactions between users.
- **Next User Forum, early 2008.**



- **gLite provides a reliable base for basic grid services. Core functionality must be augmented with higher-level software/services to provide complete stack for real applications.**
- **Continue with “cluster” development**
 - Proven model to provide tight link with user requirements
 - Focus on “general”, high-level tools that can benefit others
 - Rigorously test core and application-level services
 - List of “disciplines”: HEP, LS, ES, CC, Fusion, A&A, GO
- **RESPECT (Recommended External Software for EGEE CommuniTies)**
 - Mechanism to highlight useful products that will work with gLite
 - Ensure “external” software provides support to user community

- **NA4 activity will:**
 - Work to coordinate support services within project
 - Use regional contacts to liaise with other projects
 - Ensure that all users' needs are met
- **Focus on making EGEE a collaborative platform**
 - New, better, faster science!

- **Scientists use the EGEE grid:**
 - Routinely and heavily to speed and to enhance their analyses,
 - To share resources, and
 - To collaborate effectively.
- **Evolution of NA4 activity:**
 - Provide “support services” to full EGEE user community.
 - Build strong, self-reliant user community.
- **“Plenary” Sessions:**
 - Workflow, Uniform Grid Access, Grid Obs., Data Mgt.
- **“Parallel” Sessions:**
 - A&A, Grid Obs., Medical Imaging
- **Demo and Poster Sessions**

- **NA4 web site:**
 - <http://egeena4.lal.in2p3.fr/>
 - First point of contact for both new and existing users.
- **gLite documentation:**
 - <http://glite.web.cern.ch/glite/documentation/userguide.asp>
 - <http://glite.web.cern.ch/glite/documentation/default.asp>
 - Documentation for core middleware functionality.
- **UIG “Use Cases”**
 - http://egee-uig.web.cern.ch/egee-uig/production_pages/UIGindex.htm
 - Simple HOWTOs for common tasks
- **NA3 Training material**
 - <http://www.egee.nesc.ac.uk/trgmat/index.html>
 - Comprehensive catalog of training materials.