Science on the grid: the other user communities

Conseil Scientif que de l'Institut des Grilles
Paris, June 1st, 2010

J. Montagnat (CNRS / UNS)

With contributions from:

C. Loomis (CNRS / LAL)

M. Petitdidier (CNRS / IPSL)

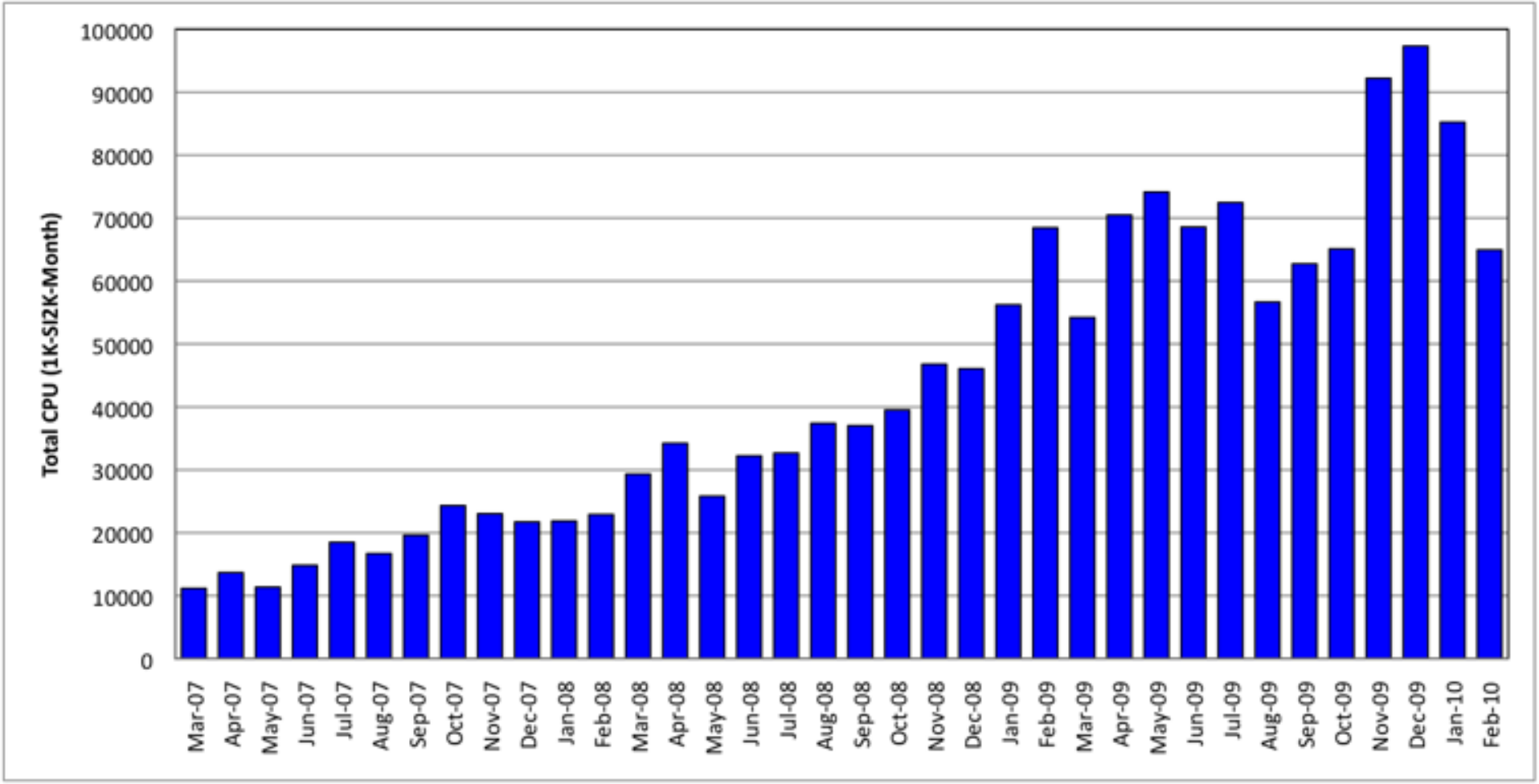
T. Glatard (CNRS / INSA / U. Lyon)

C. Blanchet (CNRS / IBCP)

V. Breton (CNRS / LPC)

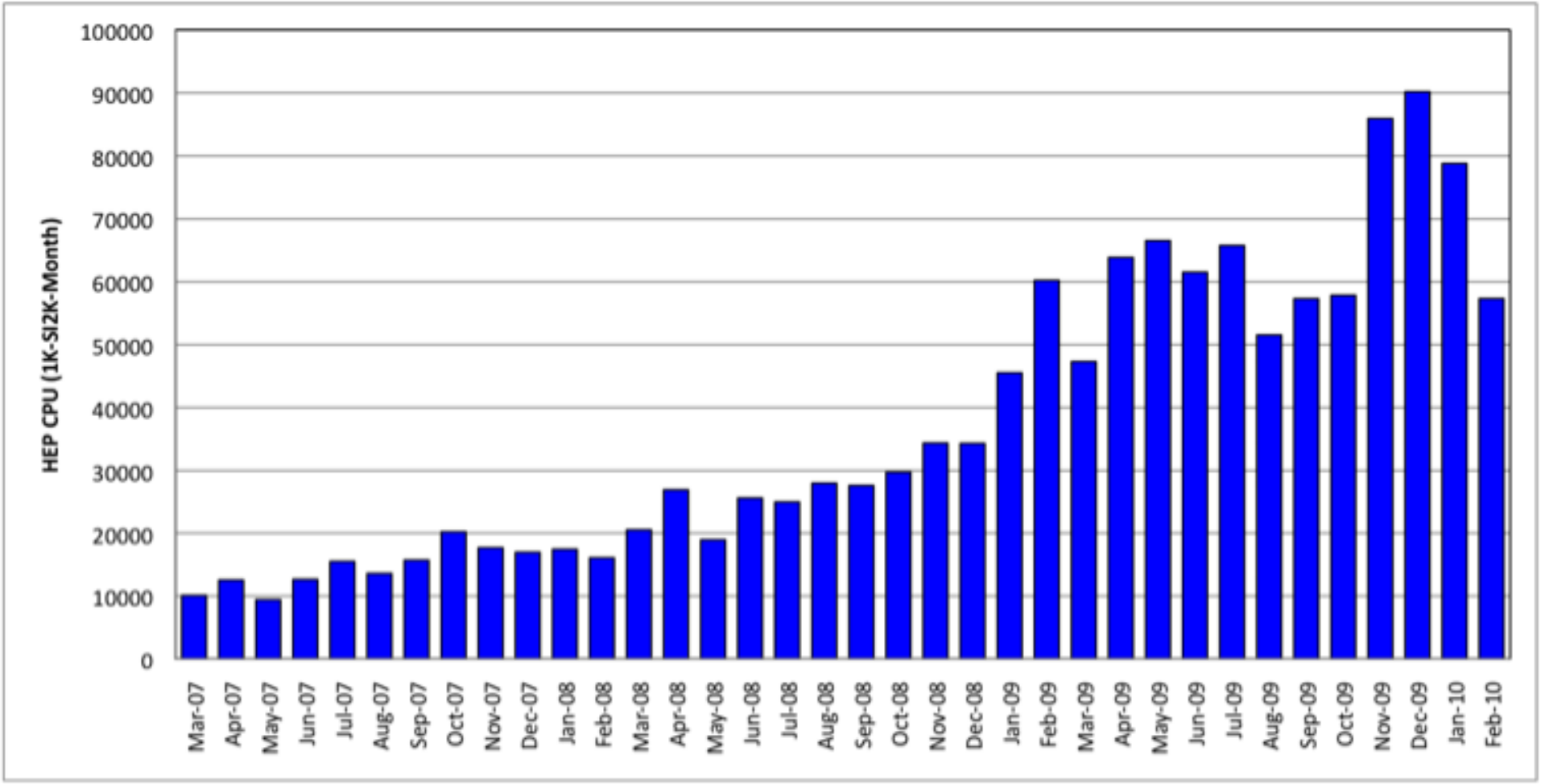
Total EGEE utilization...

- CPU utilization (normalized CPU months)

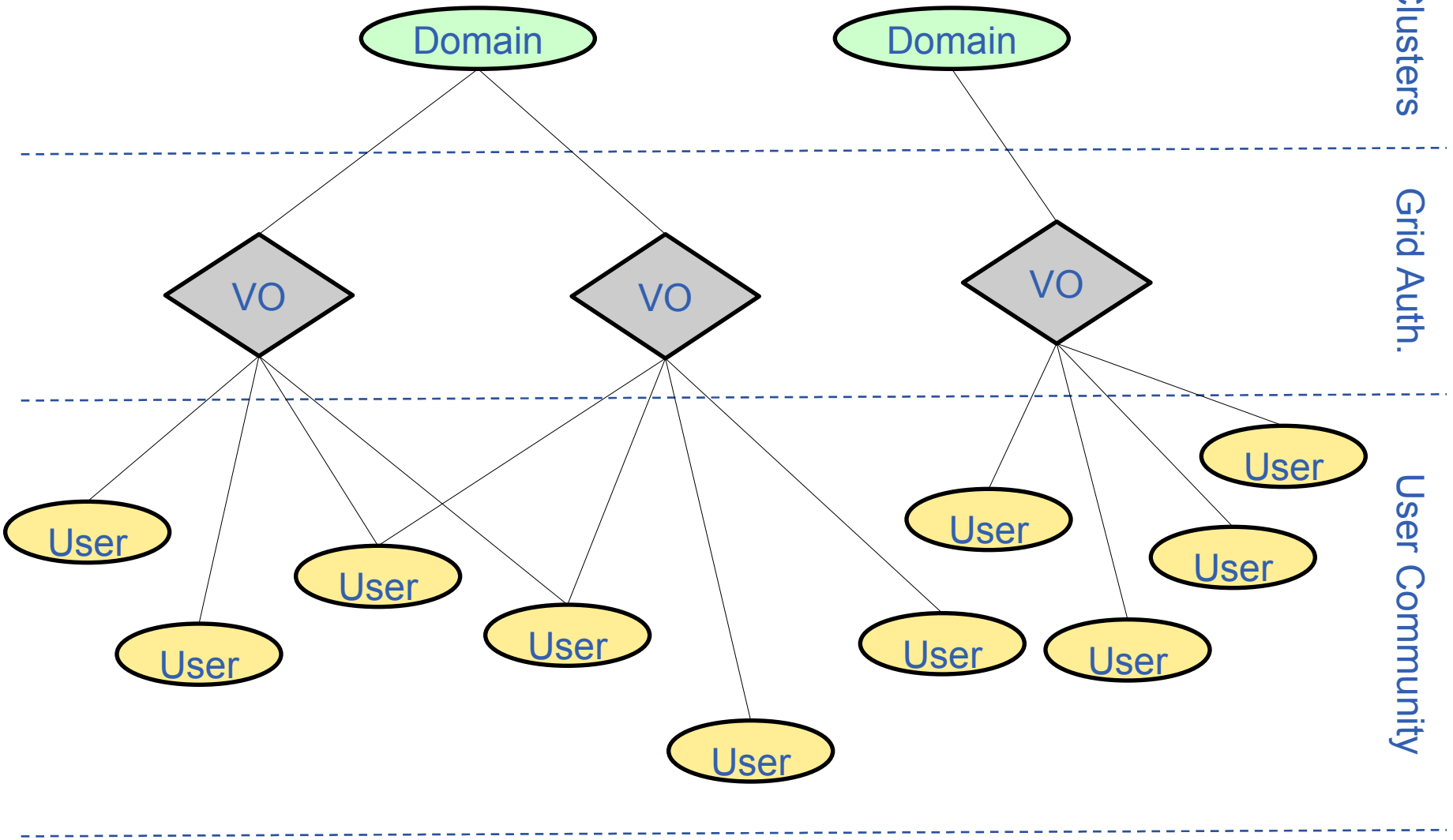


...out of which HEP...

- **Largest user community (> 92%)**
  - 4 LHC VOs (ALICE, ATLAS, CMS, LHCb)
  - Other VOs (D0, Babar...)



… and the other communities



Clusters and VOs

- **Strategic Discipline Clusters**
  - Astronomy & Astrophysics (AA)
  - Computational Chemistry (CC)
  - Earth Sciences (ES)
  - Fusion (F)
  - Grid Observatory (GO)
  - Life Sciences (LS)
- **Virtual Organization**
  - 191 VOs
  - Domain-specific VOs
    - e.g. core (legacy) VOs for strategic clusters
  - Multidisciplinary VOs (MV)
  - Regional VOs
  - Test VOs
  - … there are unknowns

Community & Use

- 191 visible VOs
- 29 “core” VOs
- Around 12500 registered users

Domain	VOs (>0%)	VOs (>10%)	VO Names (>10%)
AA	13	5	argo, auger, glast.org, icecube, virgo
CC	4	2	compchem, trgrida
CS	2	1	Vo.mosfet.es-ngi.eu
ES	5	2	egeode, esr
F	1	1	fusion
HEP	34	4	alice, atlas, cms, lhcb
INF	18	4	hungrid, nw_ru, ops, vo.rhone-alpes.idgrilles.fr
LS	8	3	biomed, enmr.eu, lsgrid
MV	15	3	balticgrid, see, vo.northgrid.ac.uk
OTH	23	1	theophys
UNK	52	3	litgrid, vo.nanocmos.ac.uk, vo.ssp.ac.uk

CIC Portal: http://cic.gridops.org/

Accounting Portal: http://www3.egee.cesga.es/

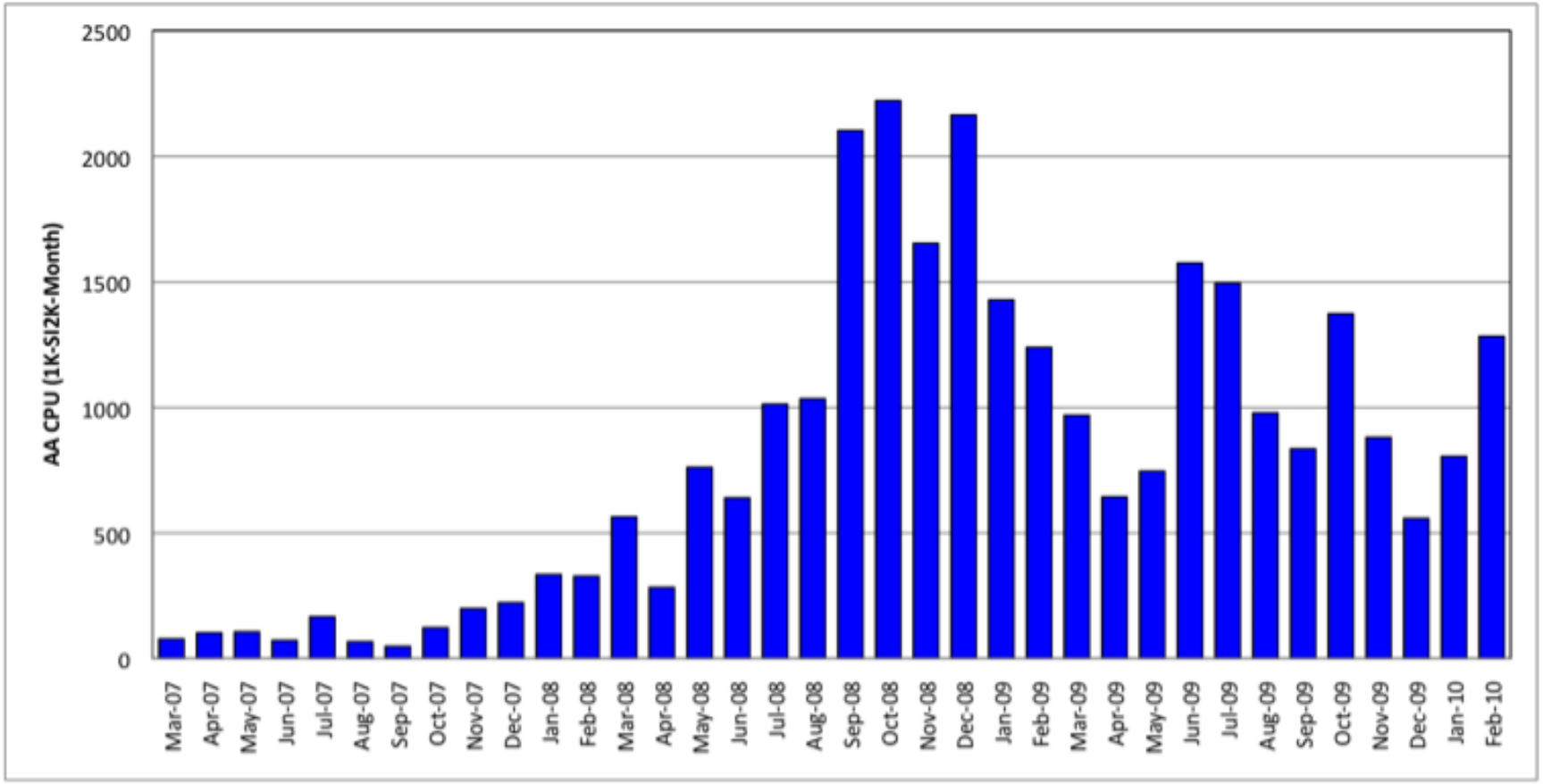
Clusters and VOs

- **Strategic Discipline Clusters**
  - Astronomy & Astrophysics (AA)
  - Computational Chemistry (CC)
  - **Earth Sciences (ES)**
  - Fusion (F)
  - Grid Observatory (GO)
  - **Life Sciences (LS)**
- **Virtual Organization**
  - 191 VOs
  - Domain-specific VOs
    - e.g. core (legacy) VOs for strategic clusters
  - Multidisciplinary VOs (MV)
  - Regional VOs
  - Test VOs
  - … there are unknowns

CNRS involved in coordination

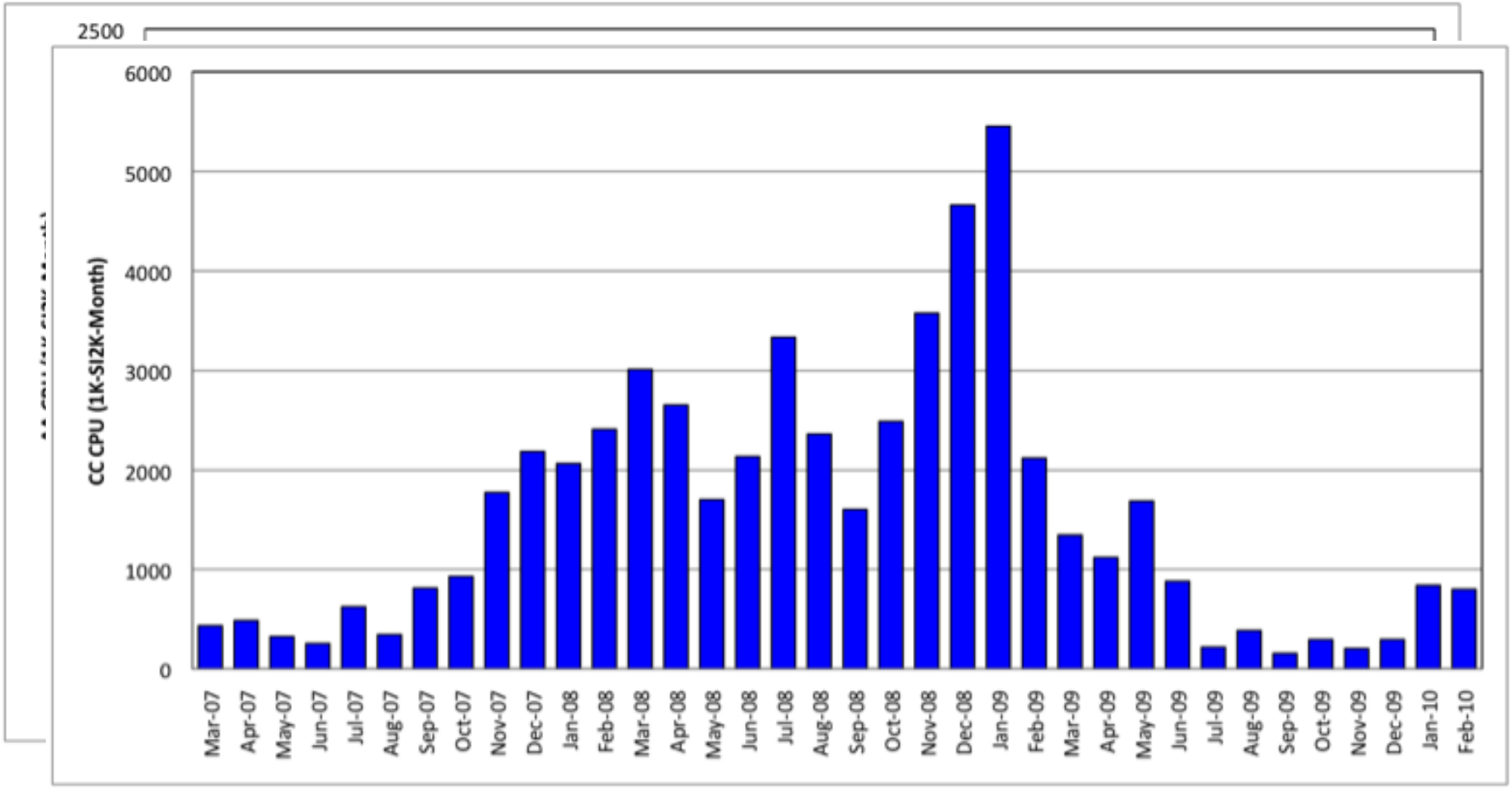
Astronomy & Astrophysics

- **New and active community**



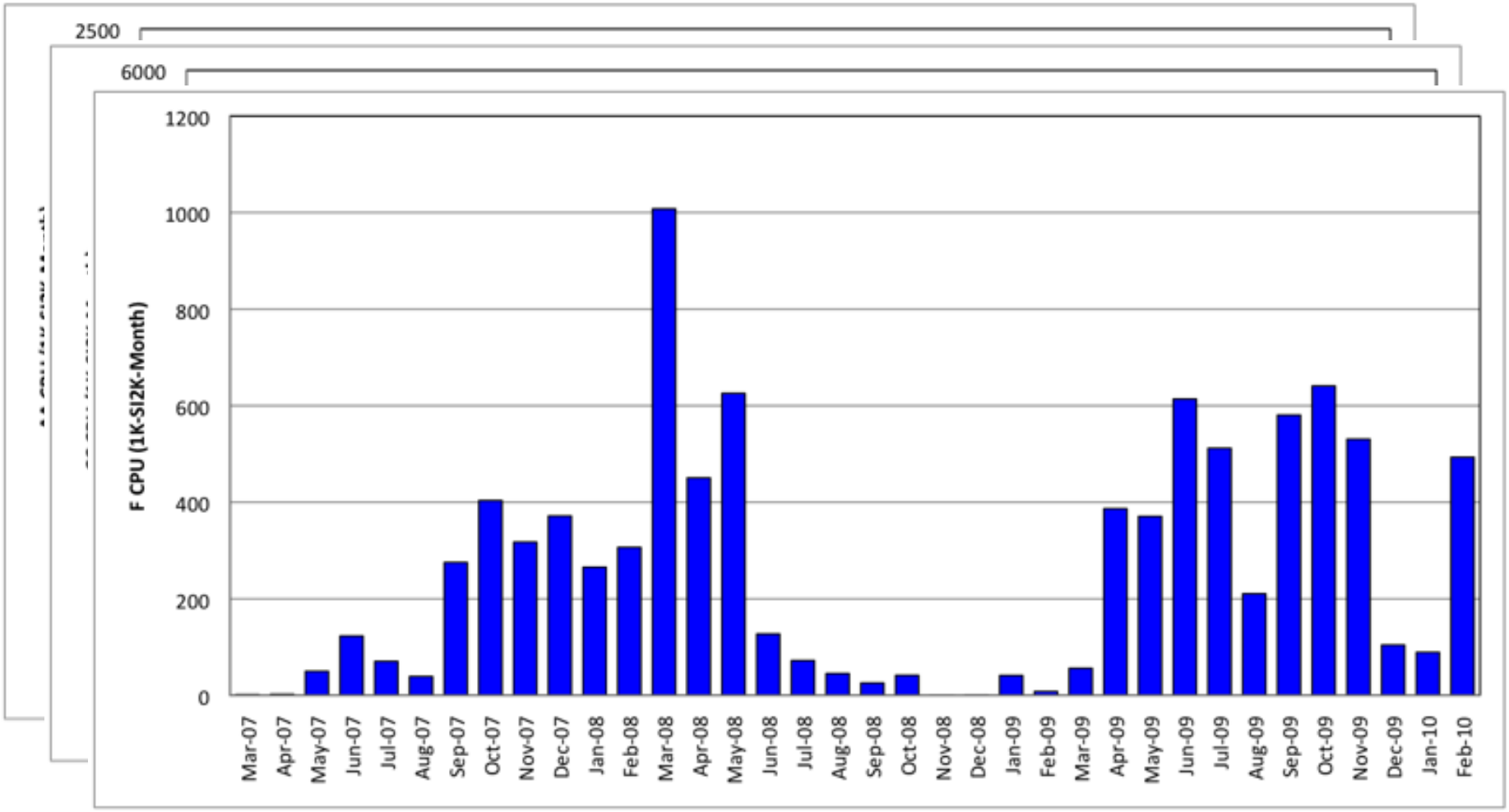
Computational Chemistry

- Has become the 2nd community in terms of resources usage



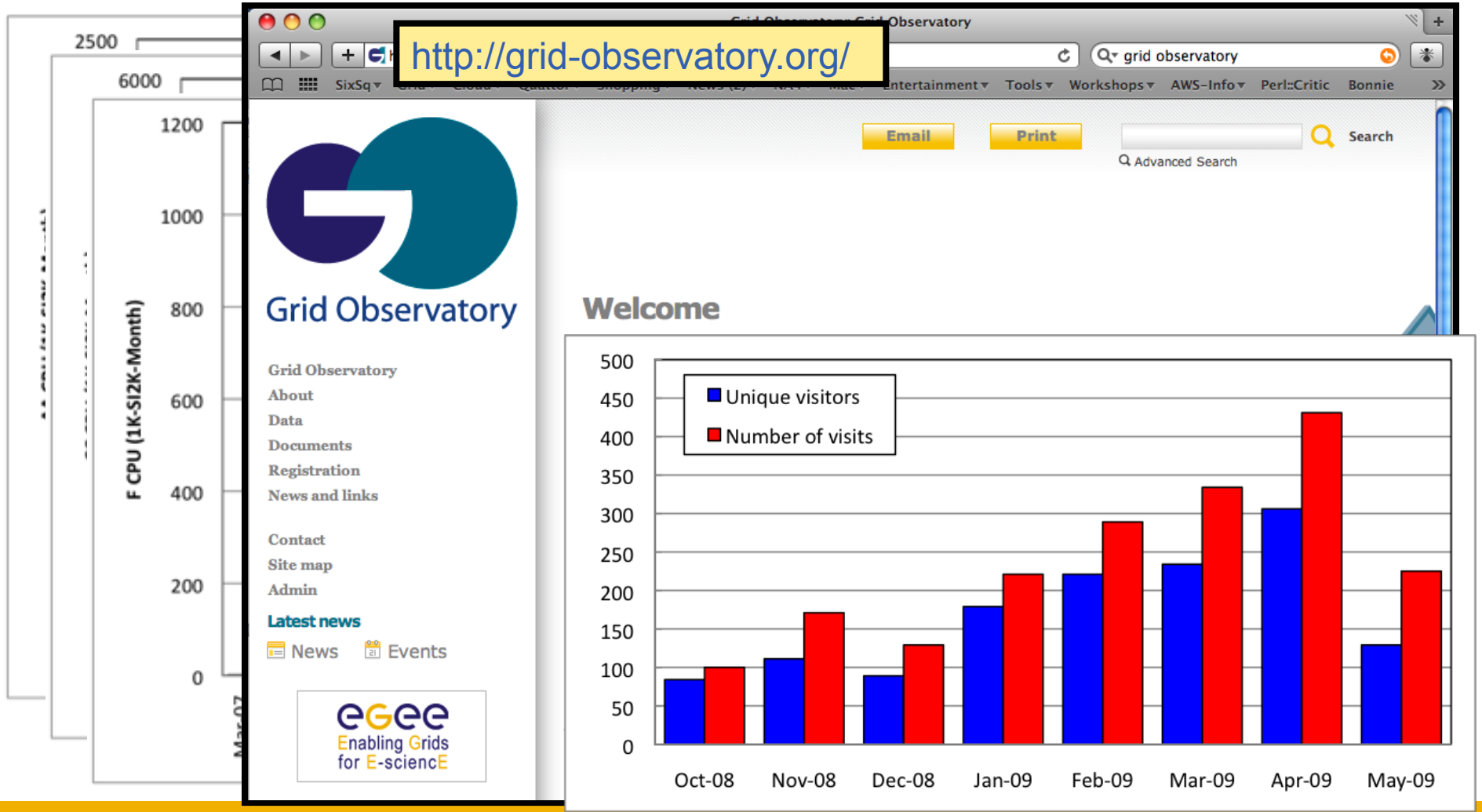
Fusion

- **Mixed computing models: distributed and super-computing**



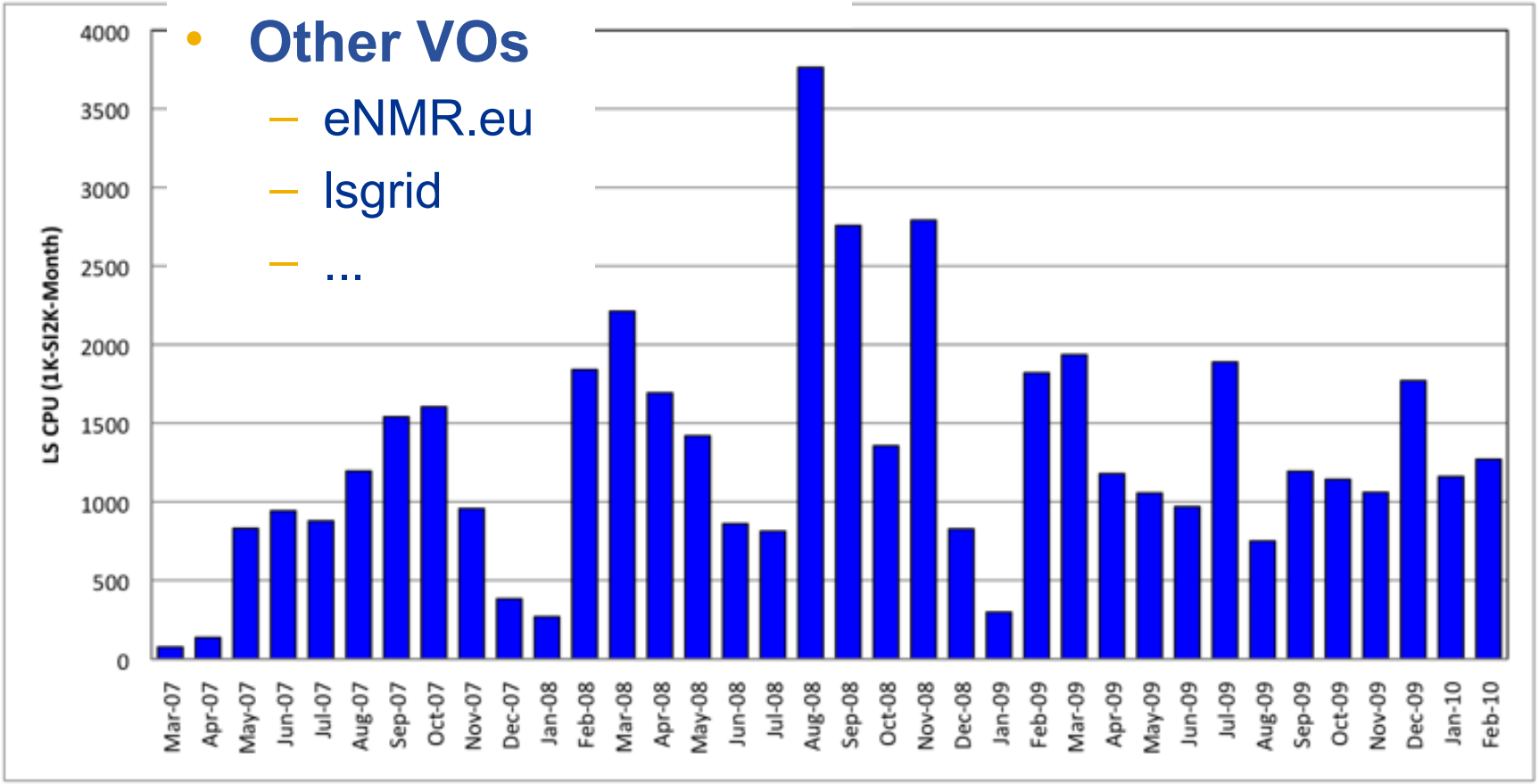
Grid Observatory

- Not production use but production usage observation
- Connection to research grids



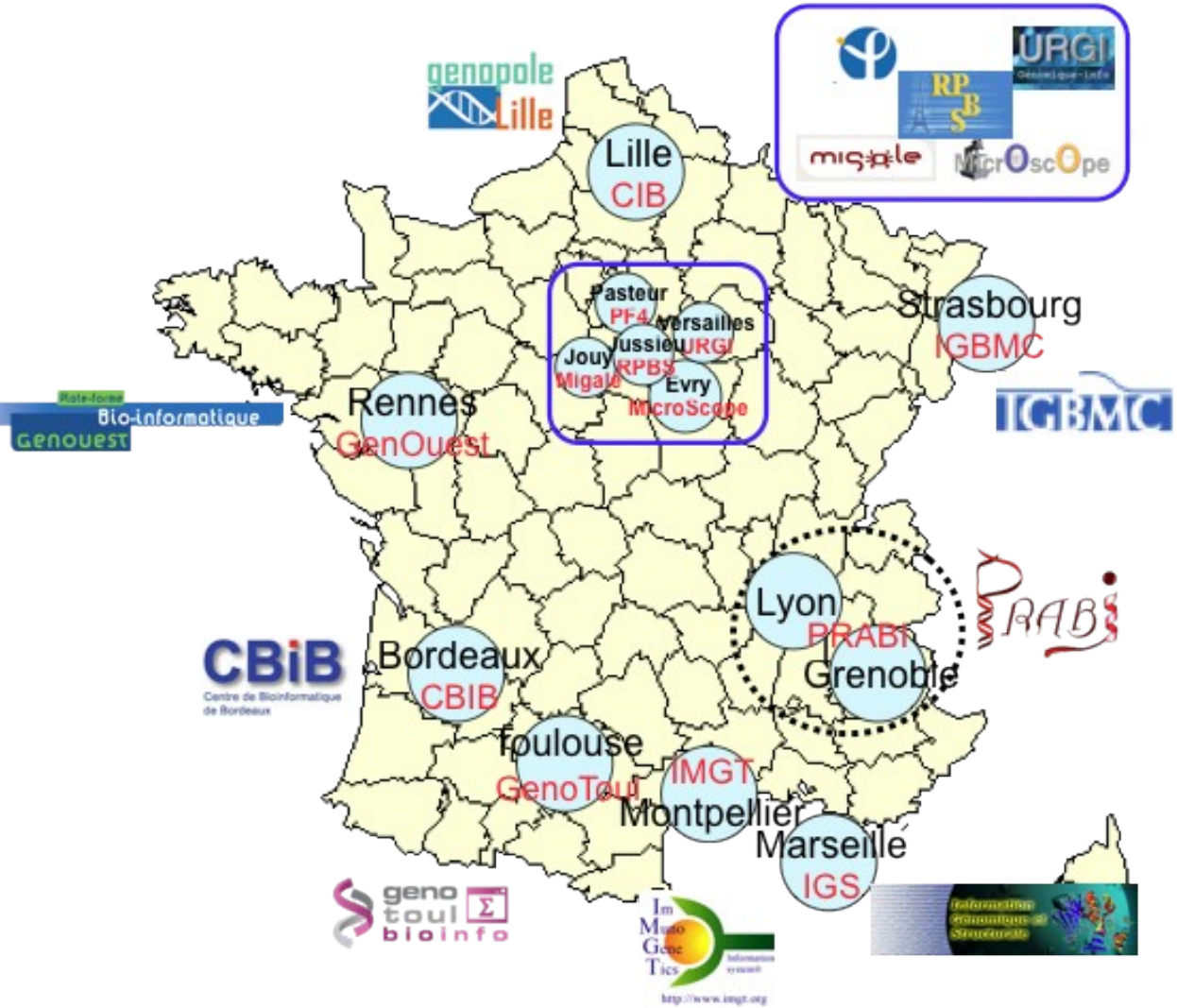
Life Sciences

- **“biomed” VO**
  - Bioinformatics
  - Medical image analysis
  - Drug discovery
- **Other VOs**
  - eNMR.eu
  - lsgrid
  - ...



RENABI

National Network of Bioinformatics Centers



Since 2004
13 national centers
(RIO/IBISA)
Bionformatics
Services to the
community

Coordinator:
Dr Claudine Medigue

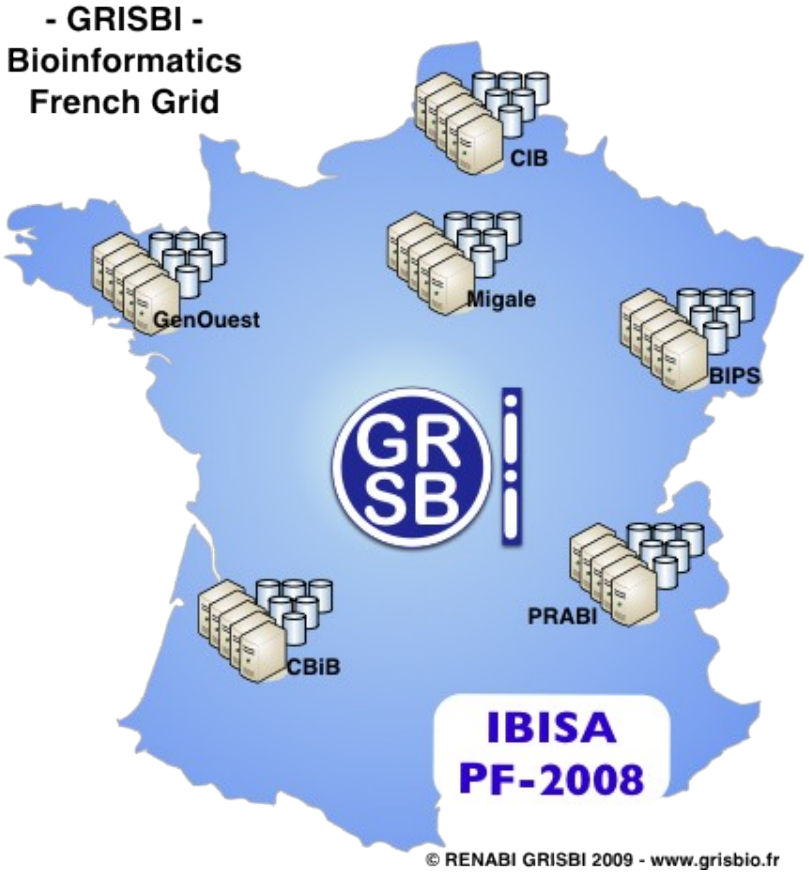
www.renabi.fr

GRISBI - Grid, Support to Bioinformatics,
www.grisbio.fr

GRISBI

Grid Support to Bioinformatics

Make possible challenging bioinformatics applications dealing with large scale biological systems



- National Production infrastructure
- 6 centers from RENABI
  - 40 participants
  - Computing resources: 1200 cores, 220 TB storage
- Applications
  - Genomics
  - Proteomics
  - Phylogeny
  - ...

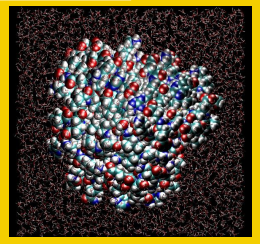
GRISBI - Grid, Support to Bioinformatics,
www.grisbio.fr

Drug discovery: WISDOM

Plateforme de Calcul pour les Sciences du Vivant

FLEXX/
AUTODOCK

AMBER

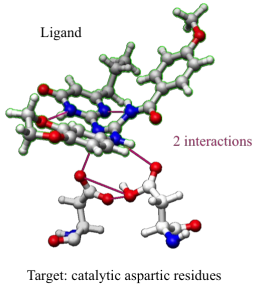


Molecular docking

Molecular dynamics

Complex visualization

CHIMERA



in vitro

WET LABORATORY

in vivo

WISDOM

Plateforme de Calcul pour les Sciences du Vivant

PCSV

WISDOM (World-wide In Silico Docking On Malaria) is an initiative aiming to demonstrate the relevance and the impact of the grid approach to address drug discovery for neglected and emerging diseases.

2005 2006 2007 2008 2009

2005	2006	2007	2008	2009
Wisdom-I	DataChallenge	**Wisdom-II**	DataChallenge	DataChallenge
Malaria	Avian Flu	Malaria	Diabetes	25M dockings (GOLD)
Plasmepsin	Neuraminidase	4 targets	Alpha-amylase	140 CPU years

Up to 400 CPU-years (over 2 months)

GRIDS

EGEE, Auvergrid, TwGrid, EELA, EuChina, EuMedGrid

EUROPEAN PROJECTS

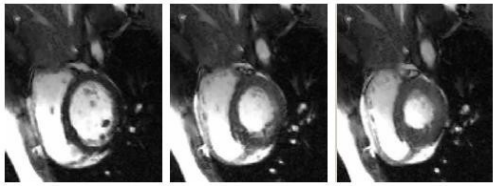
Embrace
EGEE
BioInfoGrid

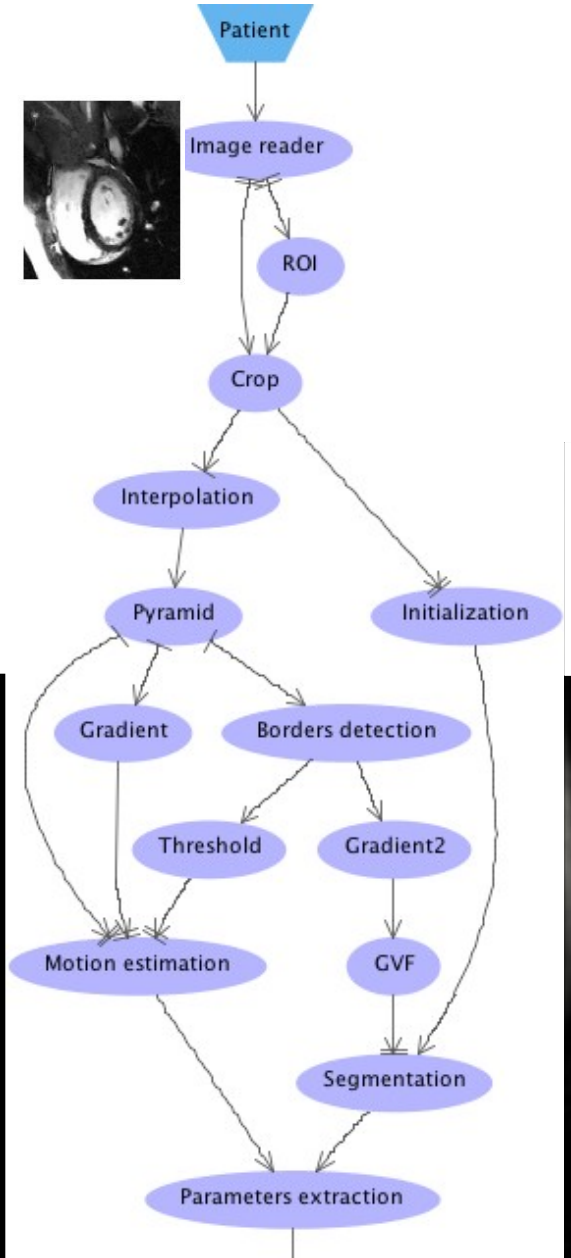
INSTITUTES

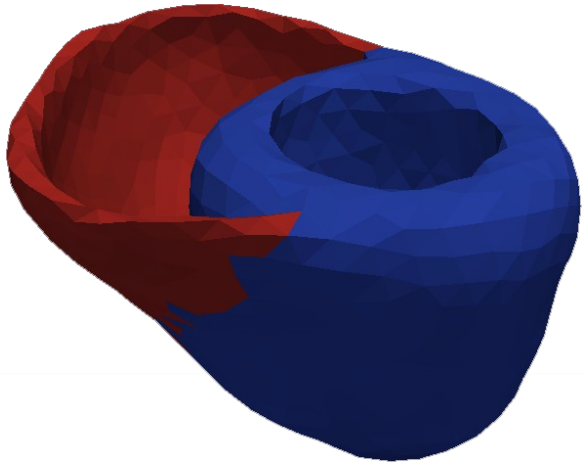
SCAI, CNU
Academica Sinica of Taiwan
ITB, Unimo Univ,, LPC, CMBA
CERN-Arda, Healthgrid, KISTI

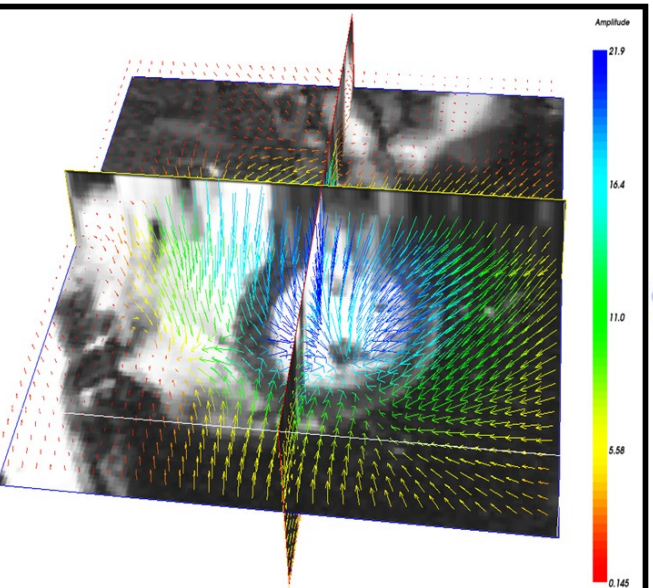
Malaria Journal 2009, 8:88 (1 May 2009)

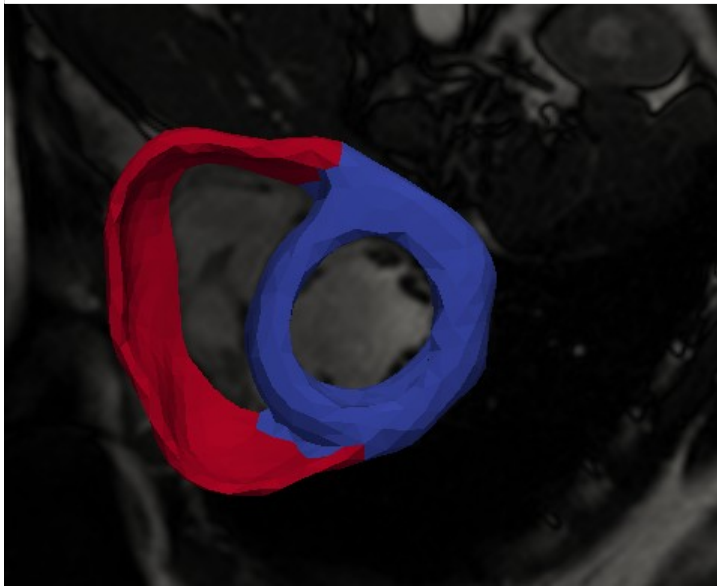
Medical Imaging: Cardiac MR sequences analysis





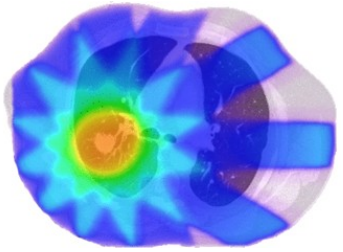


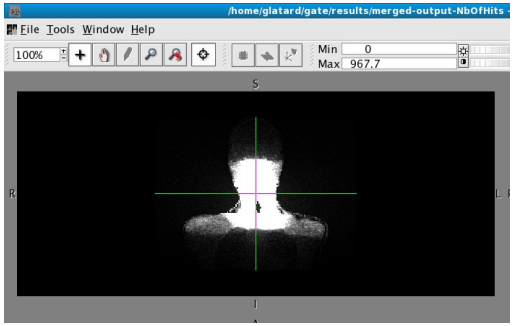




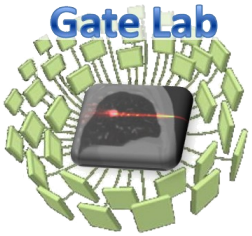
GATE-Lab

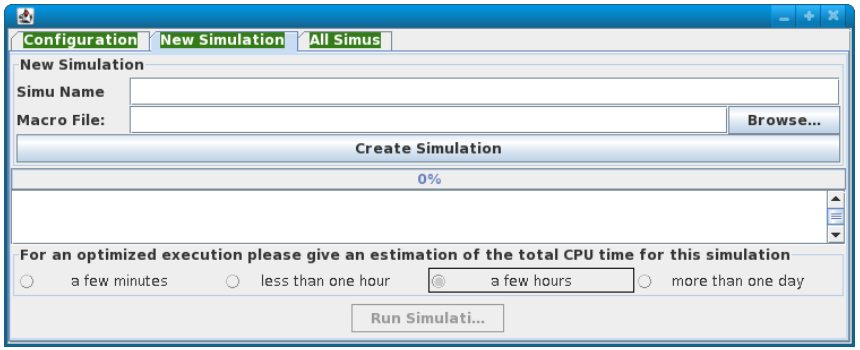
- **GATE simulations**
  - GEANT4 – based (Monte Carlo simulation)
  - Particle tracking through matter
  - Produce hit/dose maps





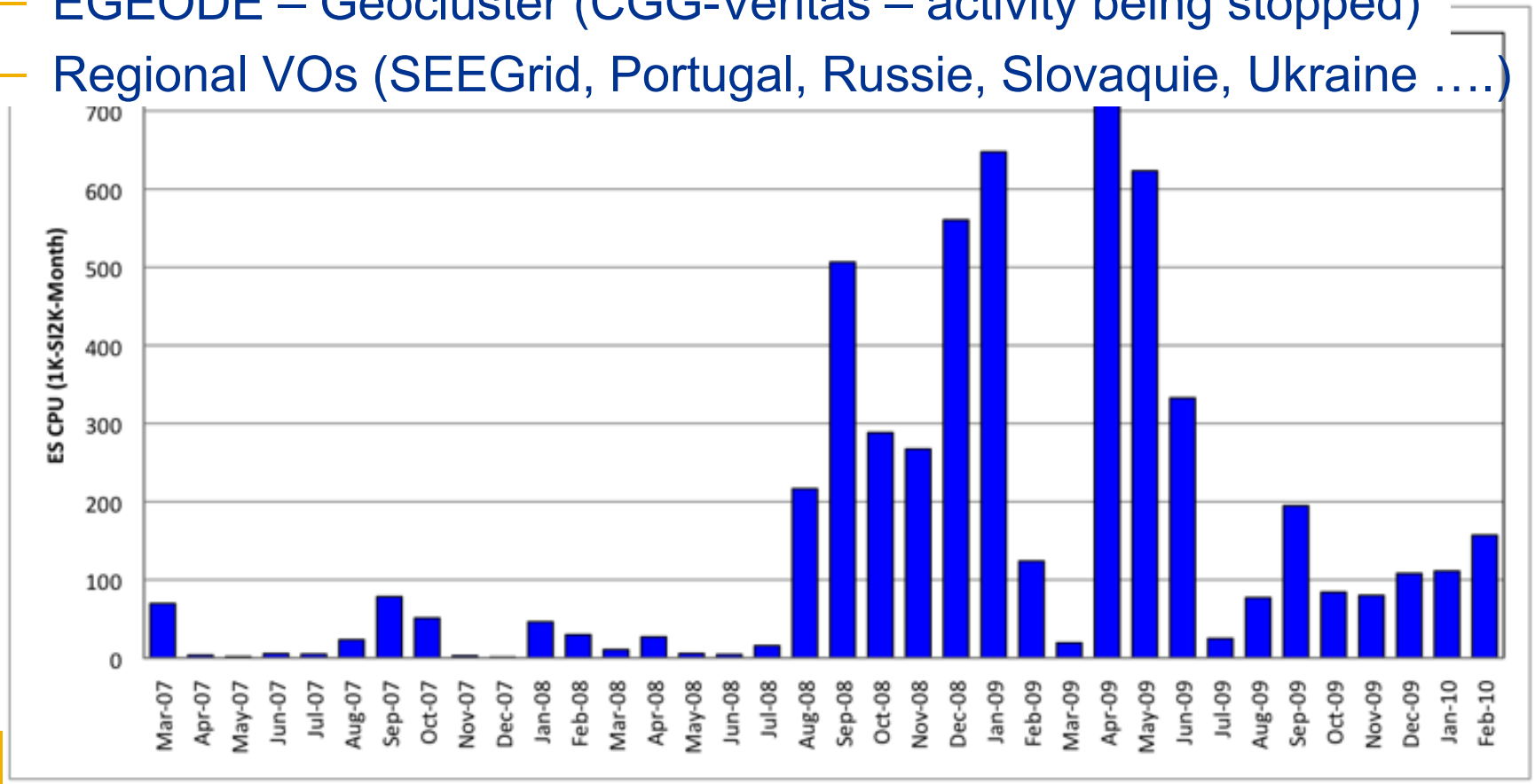
- **Multiple uses**
  - Radiology image simulation
  - Radiotheray / Hadronthearpy simulation
  - Radiology devices modeling
  - ...





Earth Sciences

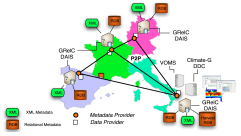
- **Multiple actors: satellite data (CNES, ESA, NASA...), meteorological data, seismic data...**
- **Virtual Organizations**
  - ESR (generic VO)
  - EGEODE – Geocluster (CGG-Veritas – activity being stopped)
  - Regional VOs (SEEGrid, Portugal, Russie, Slovaquie, Ukraine ….)



Applications & Community

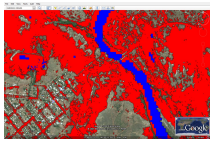
Climate - GRelC

CMCC (Italy), IPSL (France), Univ Cantabria (Spain), SCAI (Germany)



Flood

EMA (France), IISAS (Slovakia), SRI (Ukraine)



Flood map for Normanton area, Queensland, Australia derived from RADARSAT-2 14-17 Feb 2009

(SRI)

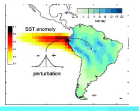
Biodiversity

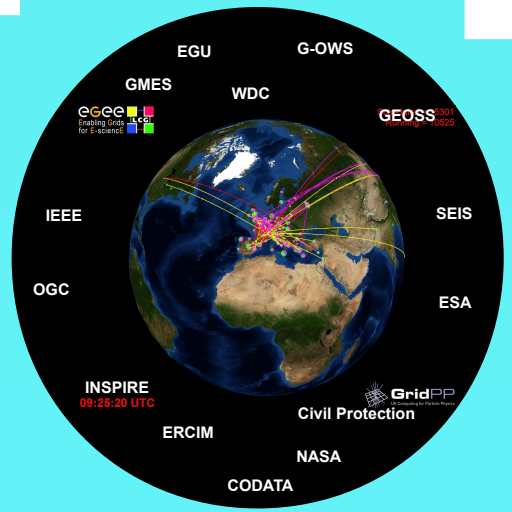
BRGM (France), Footways(France),JKI (Germany)

Scenarios climatological

Climate - El Niño

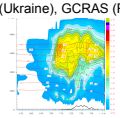
Univ. Cantabria – EELA2





Meteorology

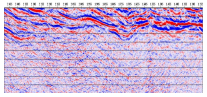
Greece: AUTH, IASA, NOA; SRI (Ukraine), GCRAS (Russia)



From IASA

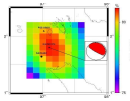
Geosciences- Geocluster

CGGVeritas (France), DSI-IRD, Geoazur, IPGP, IPGS, ISTEP, Sisyphe, UBO



Seismology

AUTH (Greece), IPGP (France), AUTH (Greece), Tubitak Ulakbim (Turkey), Univ. patras (Greece), INFP (Romania)

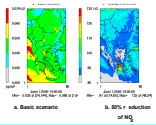


From

E. Clévédé - IPGP

Pollution

IPP-BAS (Bulgaria), IASA (Greece), EnvVO-SEEGRID

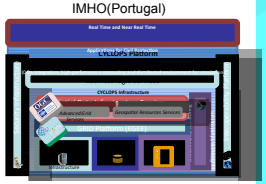


a. Basic scenario

b. 50% r eduction of NO_x

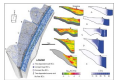
Geospatial platform

IMAA(Italy), INFN(Italy), EMA (France), IMHO(Portugal)



Hydrology

CRS4 (Italy), Univ. Genève, Univ. Neuchâtel,INHGA (Romania), Institute for Water Resources "Jaroslav Cerni", Belgrade and CSASA at University of Kragujevac, Serbia



Forecasting seawater intrusion in a coastal aquifer

From J. Kerrou

Univ. Neuchâtel

- Coordination: H. Schwichtenberg (SCAI/FhG) & M. Petitdidier (IPSL/LATMOS)
- Stakeholders in 25 countries
- Many different activities
- Well identified common interests

Results

- Biodiversity database (Footprint) BRGM
- GRelC distributed database (climate and seismology data)
- Meteorology models (weather prediction, pollution, climate)
- Sattelite data interface (GENESI-DR)
- Geospatial components used for flash flooding predictions
- Workflows (KwfGrid)
- Scientific publications
  - 20 journal papers
  - Special issue of “Grid computing in Geosciences”
  - 6 PhD thesis in France

Application communities transition towards EGI

- **Overall context**
  - Complete model change ("centralized" EU project → network of NGIs)
  - Unfunded application support EU projects (ROSCOE, SAFE...)
  - Little European-wide coordination
    - (some exceptions: CERN for HEP, EBI for bioinformatics...)
- **Needs**
  - Scientific and technical coordination
  - User support
  - Training and induction
- **Risks**
  - Adoption: grid computing is a tool, not a necessity
  - Fragmentation / duplication of efforts
  - Loose coordination / difficult interaction between VOs
  - Lack of international coordination

Transition towards EGI

- **Difficulties**
  - Find sustainable human resources to coordinate the community and operate the VO(s)
  - Scientific coverage of “domains”, number of VOs / groups...
  - Legitimacy of (national) coordinators
- **The case of Life Sciences community**
  - Current management of the “biomed” VO
    - Scientific and technical coordination was expected to come from ROSCOE proposal → “volunteer” resources
    - User support → some dedicated services support within EGI SA3, “volunteer” technical team (technical shifts organized, VO resources monitoring)
    - Training and induction → no resources
  - Other VOs in the LS domain are regional / project specific
  - No real overall coordination of the domain

LS Transition towards EGI

- **Discussions with NGIs with strong involvement in LS**
  - French, Dutch, Spanish and Italian NGIs
  - Proposal to leverage HealthGrid association. Dedicated workshop at the HealthGrid conference in Paris, end of June.
- **Actions in progress**
  - One large “biomed” VO, split in sub-activity groups
    - Identify scientific contacts for sub-activities
    - Technical constraints related to group-wise management of users
  - Better tracking of users and activity (curation of the VO, documentation of applications and their life-cycles...)
  - Technical team (technical contact, interface with operations, active problems identification role)
    - Technical shifts
    - VO-level infrastructure monitoring tools

Conclusions

- **Critical transition phase**
  - Lack of sustainable support for application domains
  - Grid far from universally adopted
- **Operational model for scientific communities still to be found**
  - Evolution of the model within EGEE-III; Further evolution needed to adapt to EGI
  - One model being implemented: still to be validated by the overall community
- **Fragmentation into NGIs difficult to handle**
  - Loose coordination cannot be avoided
  - Fragmentation of resources and communities would be the real grid killer