The Grid lanscape in France

Guy Wormser

LAL Orsay

October 10, 2006

Grids in France : present status

- **The HPC centers**
  - There are 3 main HPC centers in France: IDRIS (Orsay, CNRS), CINES (Montpellier, Ministry of Higher Education), CCRT (Bruyère le Châtel, CEA), rated at a few Tflops each
  - Very recently, the ministry decided to create a consortium with private company bylaws (« société civile ») to manage them and boost their potential
  - IDRIS is leading the European DEISA project
- **The Computing science research community**
  - Very active community present in 3 different organisms: INRIA, CNRS, Universities
  - A unique research tool : GRID5000
  - France research community is leading the EU NoE COREGRID
- **The National Research and Education Network: RENATER**
  - Very proactive to provide high speed network to the different grid nodes
- **EGEE/LCG**
  - Very active French federation in EGEE/LCG, present since the very beginning (DATAGRID)
  - Has always been in charge of key sectors : applications, network, central operations, Industry Forum
  - Around ten active nodes, with a very big one , IN2P3 computing center in Lyon
- **Actions supported by the Ministry of Research**
  - Many supported actions throughout the years thru various calls for proposals (a few M€/years), orineted mostly towards middleware development
  - ACI GRID
  - ANR « calcul intensif et grilles »
  - ANR « Masses de données »
- **Strong Involvment of some large companies**
  - Communications et Systèmes (CS)
  - Compagnie générale de Géophysique (CGG)
- **This complete grid program represents >10 M€/year**

The GRID5000 project

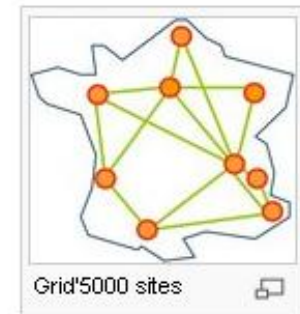- **http://www.grid5000.fr**

Grid'5000 at a glance

- **Grid'5000** project aims at building a **highly reconfigurable, controlable and monitorable experimental Grid platform** gathering **9 sites** geographically distributed in France featuring a total of 5000 CPUs:

Sites:

- Bordeaux	- Lyon	- Rennes
- Grenoble	- Nancy	- Sophia-Antipolis
- Lille	- Orsay	- Toulouse

Grid'5000 sites

- The main purpose of this platform is to serve as an experimental testbed for research in Grid Computing.
- This project is one initiative of the French ACI Grid incentive (see below: Funding Institutions) which provides a large part of Grid'5000 funding on behalf of the French Ministry of Research & Education.

- Recent increased contacts in Orsay (LAL/LRI) to port EGEE middleware on Grid 5000
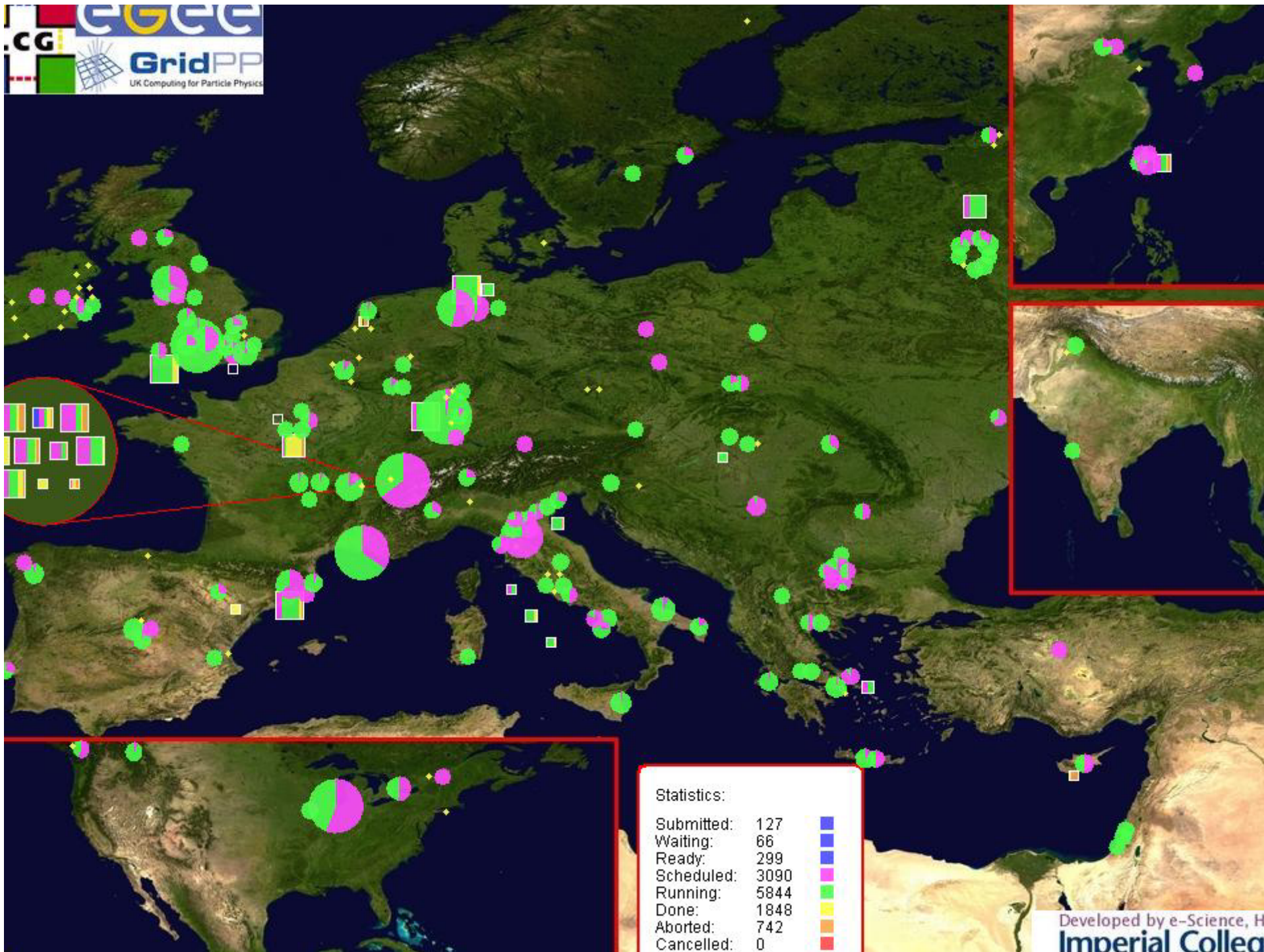
The French Federation in EGEE

- **7 partners: CNRS, CEA, Ecole Centrale Paris, CGG, CS, CNES, Healthgrid**

- **Strong contribution to the infrastructure**

- **A community of ~100 people**

- **Leaders in applications (NA4), Quality Assurance (JRA2), Networking (SA2/JRA4)**

- **Others responsabilities in EGEE:**
  - CIC team on duty
  - Industry Forum
  - Biology/Medecine Pilot applications
  - EGEE Generic Applications Advisory Panel (EGAAP)
  - External Advisory Committee (T. Priol)

EGEE Computing Resources

Resource Centers foreseen in the project

Region	CPU nodes	Disk (TB)	CPU Nodes Month 15	Disk (TB) Month 15
	Month 1: 10		**Month 15: 20**	
CERN	900	140	1800	310
UK + Ireland	100	25	2200	300
France	400	15	895	50
Italy	553	60.6	679	67.2
North	200	20	2000	50
South West	250	10	250	10
Germany + Switzerland	100	2	400	67
South East	146	7	322	14
Central Europe	385	15	730	32
Russia	50	7	152	36
Totals	3084	302	8768	936

Statistics:

Submitted:	127	
Waiting:	66	
Ready:	299	
Scheduled:	3090	
Running:	5844	
Done:	1848	
Aborted:	742	
Cancelled:	0	

The main nodes of the French Grid

- **Tier 1 : IN2P3 Computing Center in Lyon**
- **Tier 2: 3 present in the MoU**
  - Paris : large federated Tier2 implanted on 5 different close-by sites
  - Clermont-Ferrand : classical Tier2, mainly financed by rgional authorities
  - Nantes: small Tier2 dedicated to Alice
- **Independent financing for Tier 1 (central funding) and Tier2s (grab-all) (at least up to 2009)**
- **Advantages of distributed Ter2**
  - More sources of possible funding
  - Smaller costs of infrastructure (power, AC, etc..)
 - Larger manpower

- **Grids have become a great vehicle for promoting interdisciplinary contacts between HEP, many other application fields and computing scientists**
  - Bioinformatics
  - Medecine
  - Chemistry
  - Fusion science
  - Earth sciences
  - Astrophysics/astronomy
  - Neuroinformatics
  - Climate
  - Finance
  - …..
- **HEP can be proud to have been a key player in these endeavours**

Earth Observation by Satellite

Hydrology

Solid Earth Physics

Meteorology

Climate
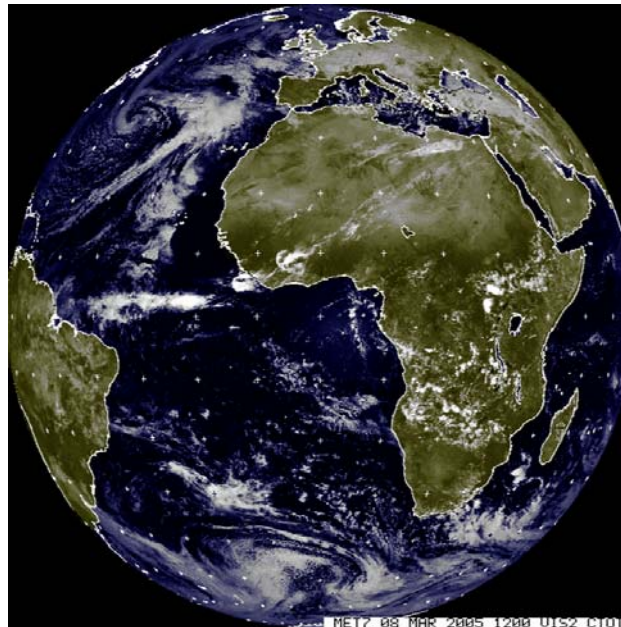
Geosciences

Chemistry of the Mars Upper Atmosphere

SEISMOLOGY[1]

Fast Determination of mechanisms of important earthquakes (IPGP: E. Clévédé, G. Patau)

Challenge

Provide results 24h -48h after its occurrence

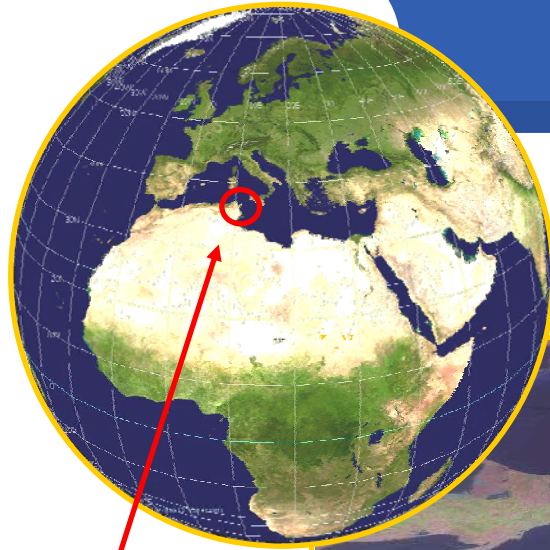5 Seisms already ported: Peru, Guadeloupe, Indonesia (Dec.), Japon, Indonesia (Feb.)

Application to run on alert

➢ Collect data of 30 seismic stations from GEOSCOPE worldwide network

➢ Select stations and data

➢ Definition of a spatial 3D grid +time

➢ Run for example 50-100jobs
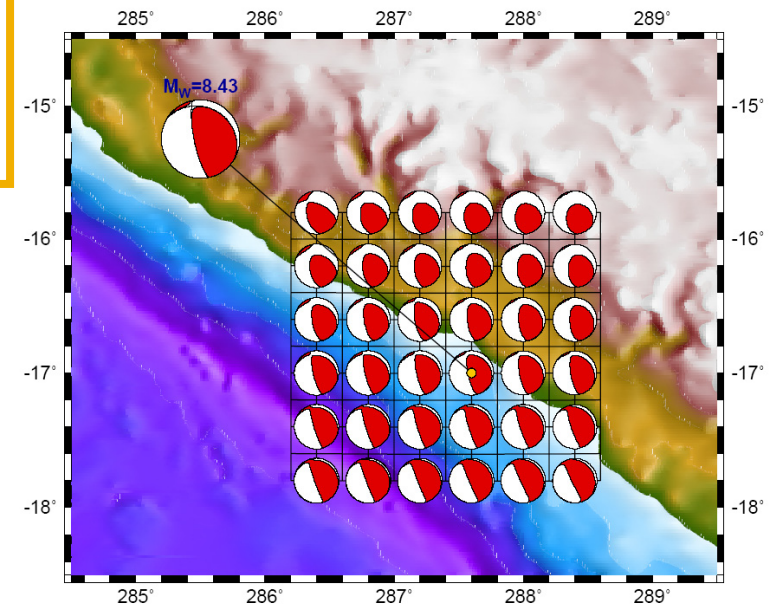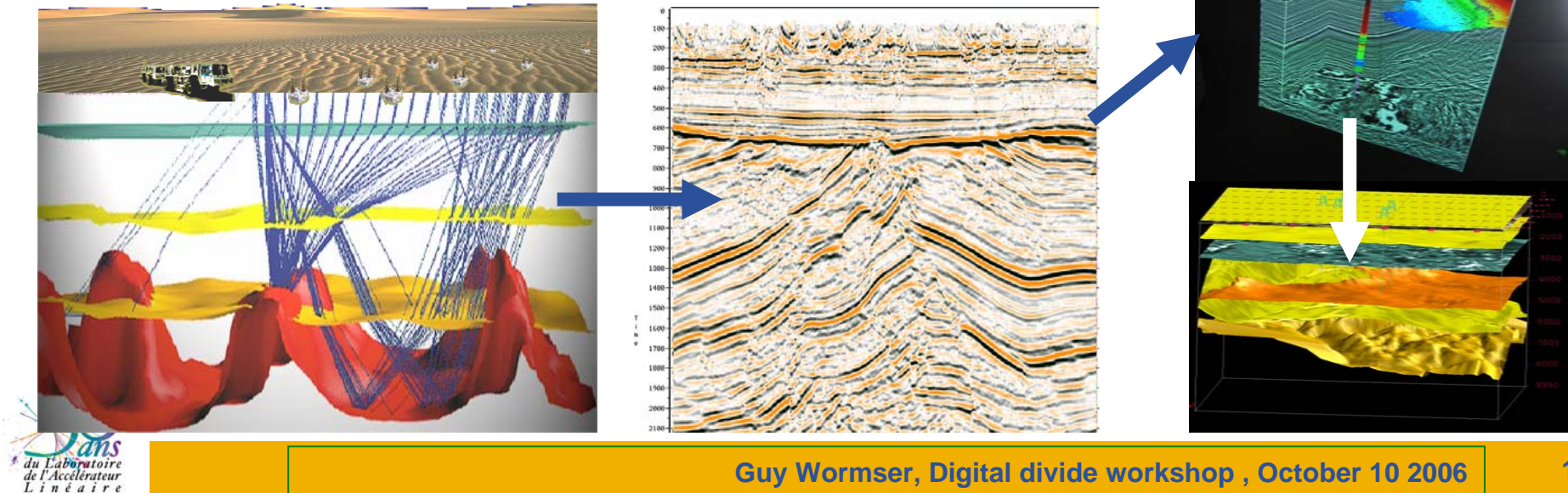

Peru earthquake, 23/6/2001, Mw=8.3

Data used: 15 Geoscope Stations

Mediterranean area (SwiMED)

G. Lecca (CRS4 Italy), P. Renard (Unine, CH),
J. Kerrou (INAT, Tunisia), R. Ababou (IMFT, Fr)

Tunisia

Cape Bon Peninsula

70km south-east of Tunis

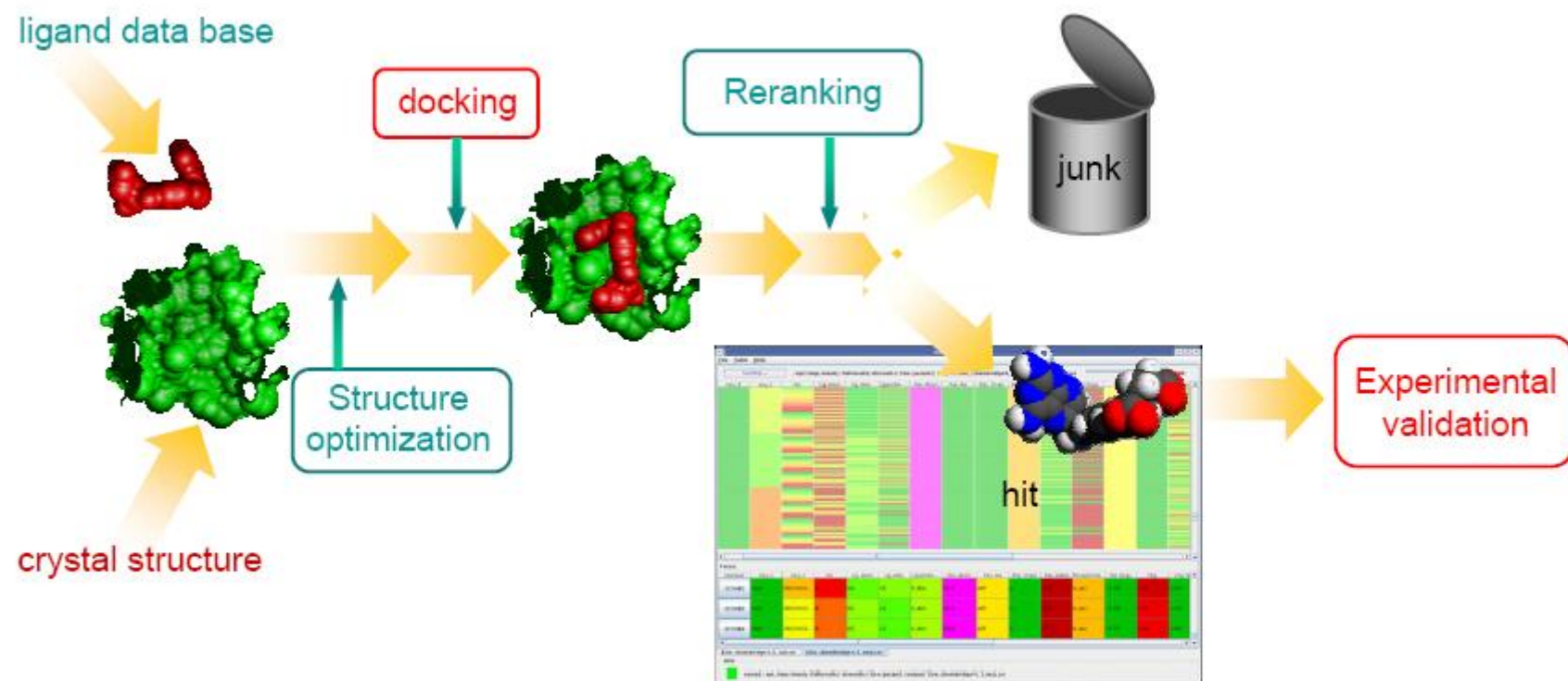**Korba coastal aquifer**

45 km

GEOSCIENCES

- **Generic seismic platform software, based on Geocluster commercial software developed by CGG**

- **Includes 400 geophysical modules, implemented on EGEE**

- **Used by both academics and private companies.**

- **Free of charge for Academics, with charge for R&D**

Status for Resource Broker IN2P3-LAL_grid09: Fri Jan 28 08:45:01 GMT 2005

RLS, VO LDAP Server: CC-IN2P3

4 RBs: CNAF, IFAE, LAPP, UPV

15 resource centres (●)
➢**17 CEs (>750 CPUs)**
➢**16 SEs**
4 RBs
1 RLS
1 LDAP Server

Dataflow and workflow in a virtual screening

Fraunhofer SCAI
Institute Algorithms
and Scientific Computing

Data challenge scenario

	Scenario
Duration	28 days
CPU time	11 years CPU
Grid performance	50%
Max number of CPU used	1,008
Number of grid jobs (20h)	12,215
Storage	2*6 TB
Docking workflow description	
Number of software / targets / compounds / parameters settings	2 / 5+3 / 500,000 / 4 = 32 mio dockings
Objective	Selection of the best hits with short analysis

FlexX running time : 1 mn
F. output size : 1MB
F. job output size : 1.2GB
F. job compressed output size : 250MB

Autodock running time : 2.5 mn
A. output size : 1MB
A. job output size : 0,5GB
A. job compressed output size : 100MB

SCAI 2005

SCAI

Fraunhofer Institute Algorithms
and Scientific Computing

page 12

Grids in France : the missing pieces

- **Many very active GRID communities but not enough contacts between them**
- **No internal structure around the EGEE project and teams**
  - Good financial support but lack of visibility and evaluation
- **Difficulty to support infrastructure activities in recent ANR calls**
- **Full financing of LCG Tier1/tier2 nodes not yet finally approved**
  - Independant schemes for Tier 1 and Tier2 financing !
- **Recent initiative within CNRS , with the support of French federations partners: The ICAD project as a nucleus for a National (EGEE style) Grid Initiative**

The ICAD proposal

- **« Programme Interdisciplinaire de Recherche »: high visibility tool with CNRS**
  - Adopted thru Conseil Scientifique and Conseil d'Administration CNRS
  - Can be supported by other institutions
  - Only 15 such programs presently running
- **Provide a distributed computing and storage infrastructure as a ressource to the scientific community**
- **Built upon the present EGEE/LCG infrastructure**
- **The main lines of action**
  - Internal structuration within CNRS and nationally
  - Central national core in the framework of a permanent European Grid infrastructure
  - Research animation with the computing science community
    - New middleware
    - EGEE Behaviour analysis
  - New applications development, both horizontally and vertically
  - Training, support and dissemination
  - Contact point with supercomputer centers
- **Budget growing from 0.5 M€ to 1 M€/year (outside salaries and materia**

The next steps

- **Formalize partnership between CNRS and RENATER**
  - RENATER ready to provide many services, beyond advanced network:
    - Certificates
    - Monitoring
    - Information services
    - Middleware distribution
    - Test bed (a la Gilda)
  - CNRS in charge of operations and applications
- **Get support from Ministry for a National Grid Initiative**
  - How to get all universities involved
  - Insert this new facility in the overall panorama
  - Good discussions already started
  - Renewed contacts in view of G8 presentation by G. de Robien in November

French position of an European grid initiative

- **Unanimous support to make EGEE/LCG success a permanent one!**

- **General consensus on the RENATER/DANTE/GEANT paradigm**

- **No detailed discussions yet on the precise structure/attributions of the central core**

- **Obvious need for homework to create the corresponding national grid initiative**
  - Sttruggle through inertia, supercomputer lobby, computer scientists lobby,etc…
  - Will take time

Conclusions

- **Many different communities very active in France on various flavours of Grids:**
  - HPC Grids (DEISA)
  - GRID5000 (unique tool for research)
  - EGEE/LCG
- **Leading roles of French teams in several key aspects of Europeans programs**
- **Major initiative underway to create the seed of a national Grid (EGEE flavour) Initiative**
  - Programme Interdisciplinaire de Recherche du CNRS, ICAD
  - Strong partnership CNRS/RENATER to be formalized
- **General support for a permanent European Grid infrastructure**

  Precise details of the core structure to be discussed