



Enabling Grids for E-science

# ***In silico* docking on EGEE infrastructure, the case of WISDOM**

***Nicolas Jacq***

***LPC of Clermont-Ferrand, CNRS/IN2P3***

***EGEE User Forum***

***CERN, 01-03.03.2006***

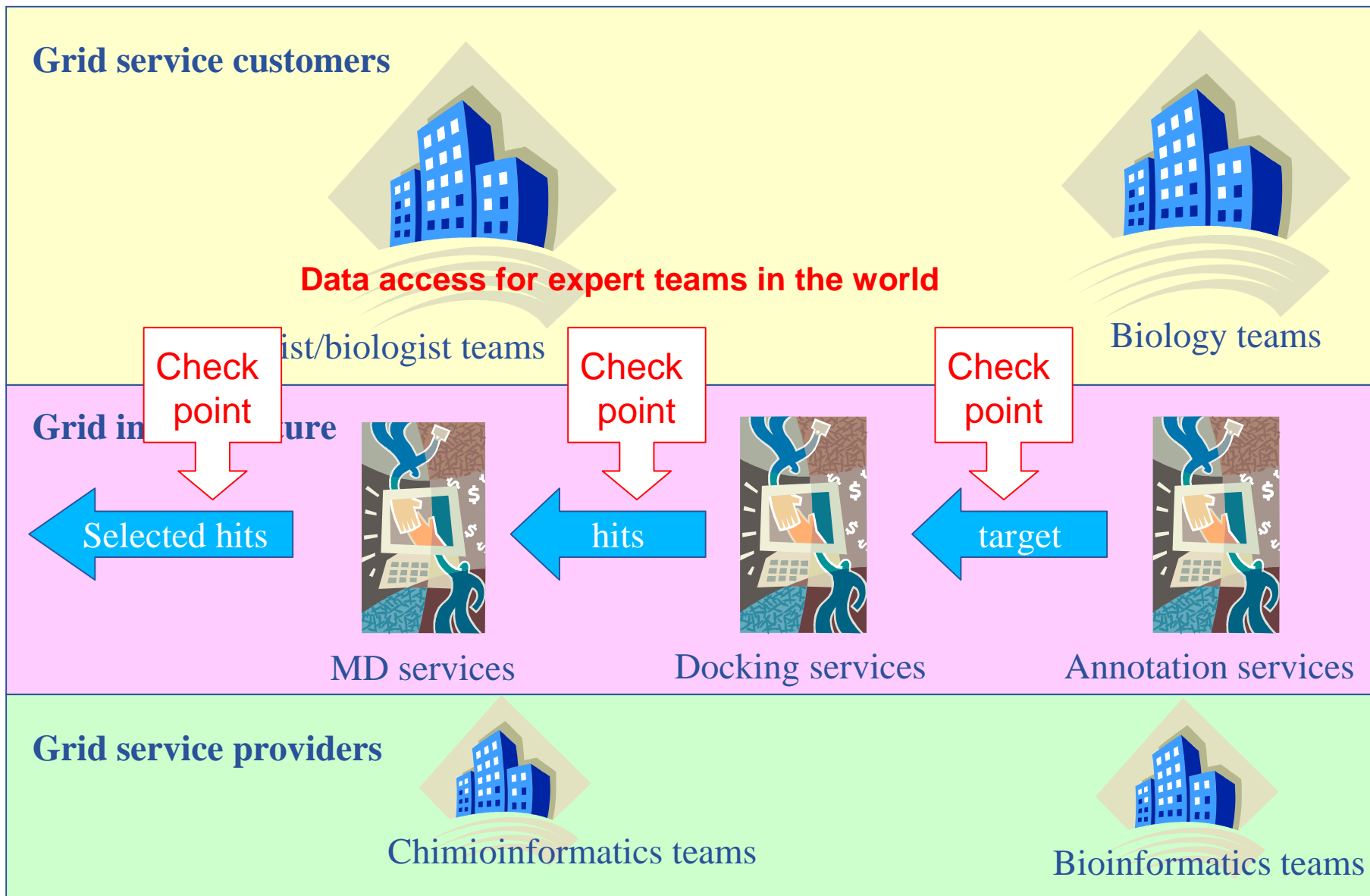
[www.eu-egee.org](http://www.eu-egee.org)



Information Society



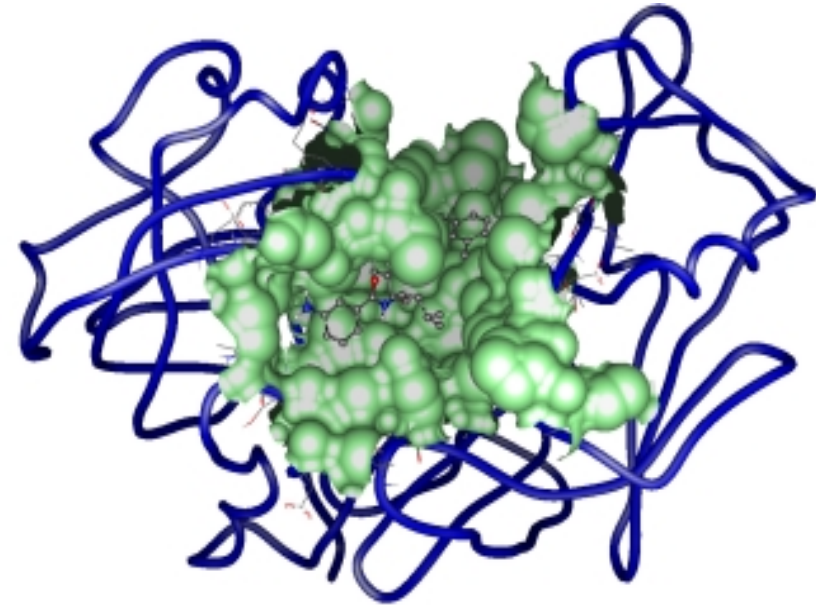
- **There is a need to develop new drugs for the diseases of the developing world**
  - HIV/AIDS, malaria and tuberculosis account for 5,6 million deaths
  - Permanent necessity to develop new drugs to fight emerging resistance to drugs (malaria)
  - Unchanged pharmacopeia for decades against trypanosomiasis, leishmaniasis, Chagas disease...
- **WHO Tropical Disease Research program is preparing a list of recommended targets for drug discovery**
- **Millions of chemical compounds are available in the laboratories and also in 2D, 3D electronic databases**
- **Set-up a world wide initiative to address *in silico* drug discovery against neglected diseases on grid infrastructures.**



- **Key issues to promote the grid in the pharmaceutical community**
  - Cost and time reduction in a drug discovery development
  - Security and data protection
  - Fault tolerant and robust services and infrastructure
  - Transparent and easy use of the interfaces
- **Grid added value of EGEE for WISDOM**
  - Large computing and storage resources
  - Job Management Service
  - Information and Monitoring Services
  - Data Management Services
  - Security (to be improved)
  - Reliability of services (to be improved)

- **Significant biological parameters**

- Two different molecular docking applications (Autodock and FlexX)
- About one million virtual ligands selected (ZINC)
- Target proteins from the parasite responsible for malaria

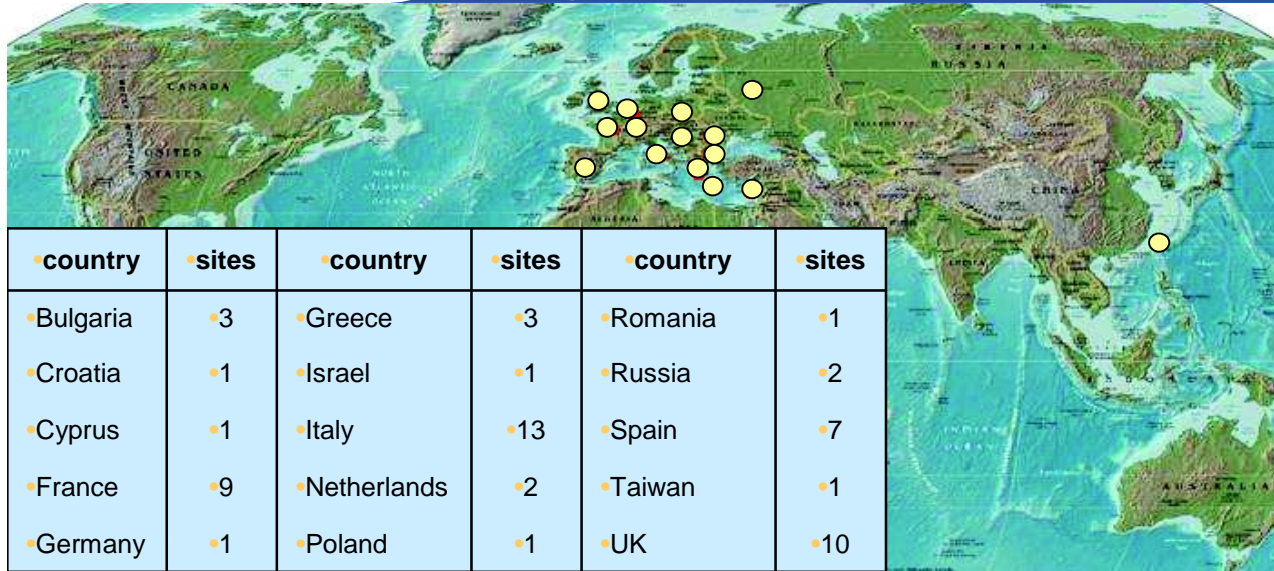


- **Significant numbers**

- Total of about 46 million ligands docked in 6 weeks
- 1TB of data produced
- Up 1700 computers in 15 countries used simultaneously corresponding to about 80 CPU years
- Average crunching factor ~600

- **Significant results**

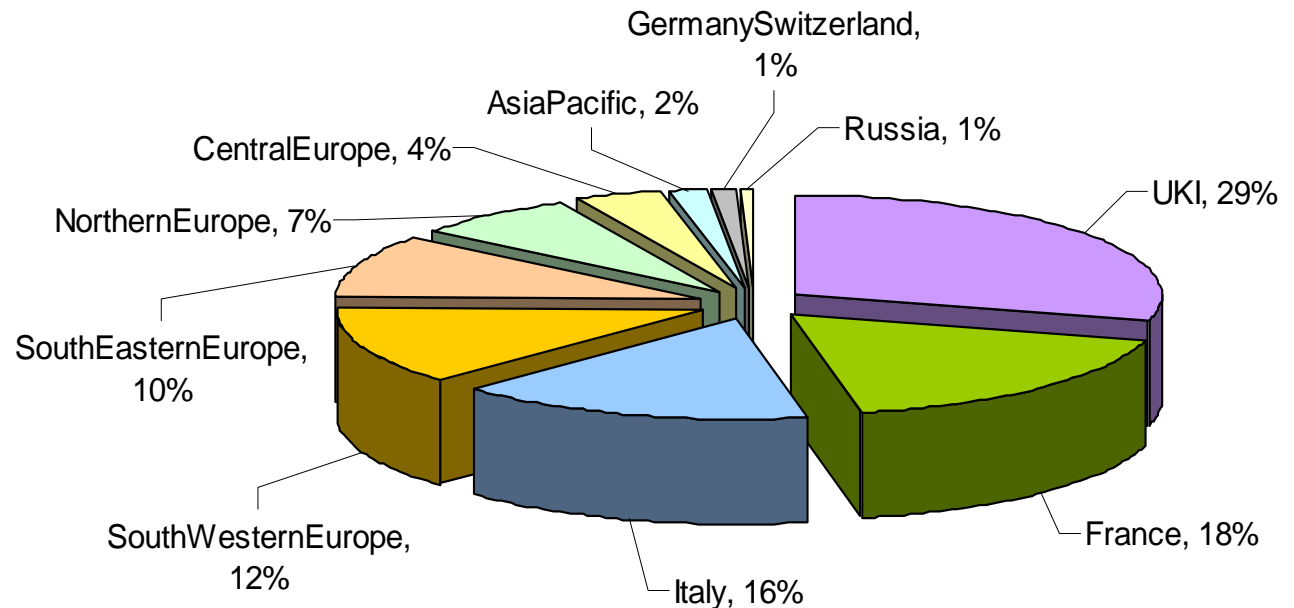
- Best hits to be reranked using Molecular Dynamics simulations

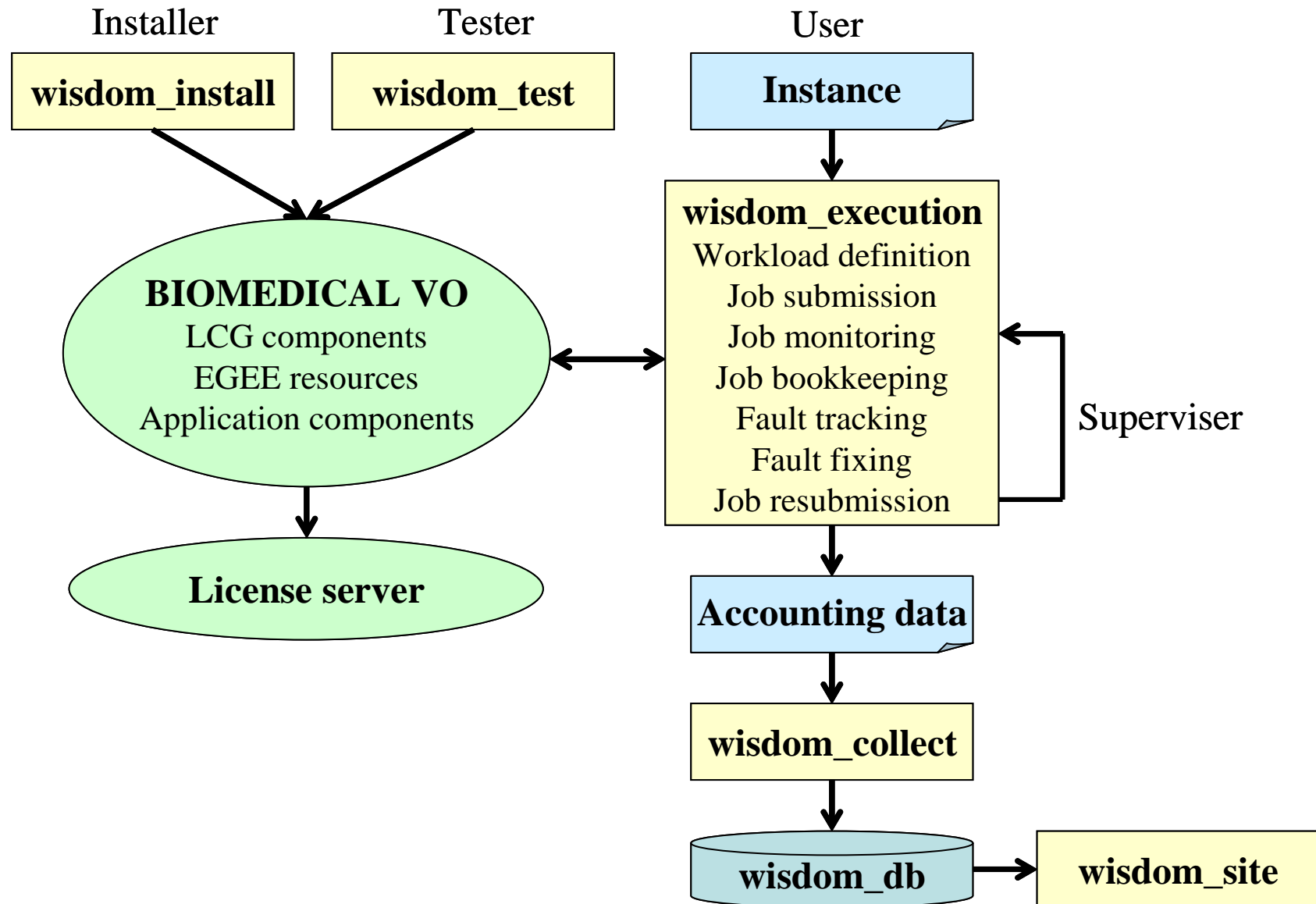


**Countries with nodes contributing to the data challenge WISDOM**

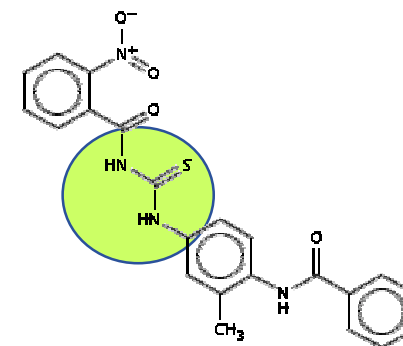
country	sites	country	sites	country	sites
Bulgaria	3	Greece	3	Romania	1
Croatia	1	Israel	1	Russia	2
Cyprus	1	Italy	13	Spain	7
France	9	Netherlands	2	Taiwan	1
Germany	1	Poland	1	UK	10

**Total amount of CPU provided by EGEE federation**

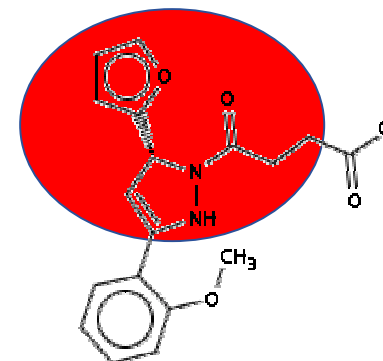




- **Conditions controlled**
  - Score of an output is independent of the grid resource where the job runs
  
- **10% compounds of Chembridge (ZINC) may be hits**
  - Top scoring compounds possess basic chemical groups like thiourea, guanidino, and amino acroleins core structure.
  - Identified compounds are non peptidic and low molecular weight compounds
  - The identified compounds look like thrombin inhibitors.



WISDOM-375228



WISDOM-113696

- **Very short term = Spring 2006 : reranking of WISDOM hits by Molecular Dynamics simulations**
  - Approximately 100 years CPU needed
  - Supported by EGEE-II & BioinfoGrid european projects
  - Need for ressources on supercomputers (contact with DEISA)
  
- **Short term = fall 2006 : WISDOM2, second large scale grid docking**
  - several new foreseen targets on malaria, dengue and other neglected diseases.
  - Resources needed: up to 80 years CPU per target
  - Supported by EGEE-II and EELA european projects, Swiss BioGrid initiative
  
- **Mid term = Summer 2007: reranking of WISDOM2 hits by MD simulations**

## LPC (CNRS/IN2P3)

- V. Breton
- N. Jacq
- J. Salzemann
- Y. Legré
- M. Reichstadt
- F. Jacq

## Fraunhofer SCAI

- M. Hofmann
- M. Zimmermann
- A. Maaß
- M. Sridhar
- K. Vinod-Kusam
- H. Schwichtenberg

## EGEE

- Biomed Task Force
- EIS team
- JRA2 team



Fraunhofer Institut  
Algorithmen und Wissen-  
schaftliches Rechnen

