Deployment of CompChem VO

Osvaldo Gervasi and Antonio Laganà

University of Perugia, Italy

Den Haag, Nov 24, 2004

Summary

- **Introduction**
- **The CompChem VO**
  - The COST in Chemistry Initiative
  - The GEMS project
- **Deployment of the VO**
  - GILDA test bed
    - The **ce.grid.unipg.it Resource Centre** (RC)
  - Implementation of GEMS prototype
  - Consortium agreement between members
  - List of members
- **Conclusions**

Introduction

- **The problem of simulating in a realistic way complex processes on a molecular basis without resorting into phenomenological or ad hoc empirical approaches is a vital need for scientists**
  - investigating new materials
  - biological systems
  - life science
  - food and drug action
  - chemical processes …
- **Up to now only specific models limited to simplified examples have been considered due to the difficulty of dealing with the complexity of realistic systems.**

Introduction (i)

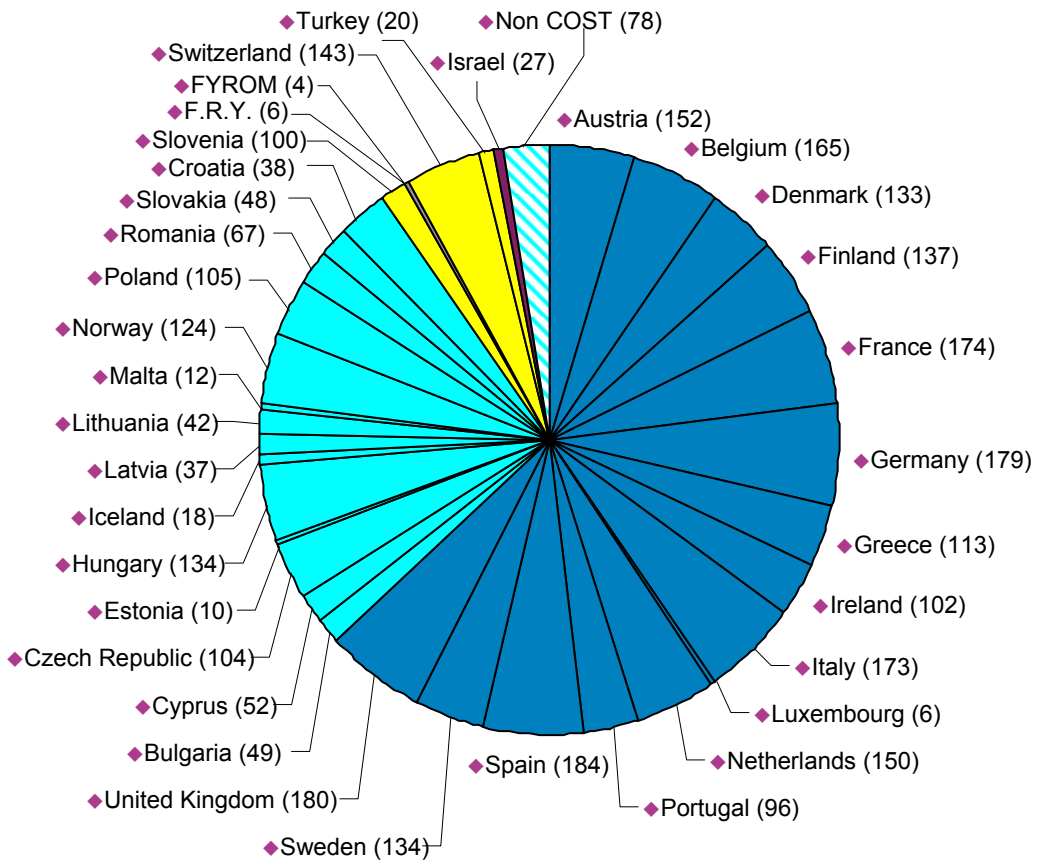
- **The rapid development of Grid technologies is making this possible by**
  - connecting the expertise, software and hardware needed to build a Molecular Simulator on a single distributed system
  - coordinating the various phases through a web based Workflow management environment
  - using a common User Interface (UI) to give to the users the access to the Grid facilities
  - enabling the usage of a large number of computer nodes
  - implementing a good security infrastructure

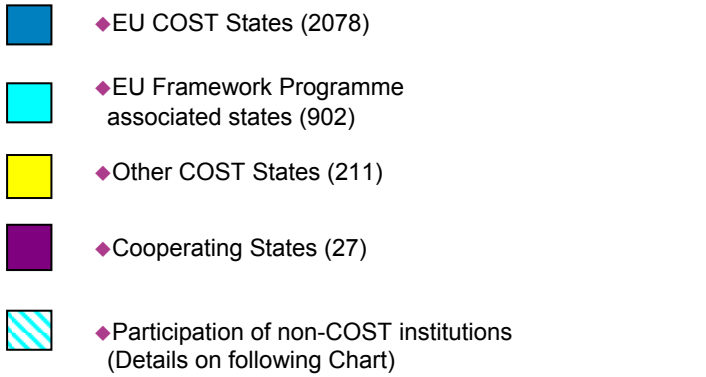
The CompChem VO

- The CompChem VO originates from the collaboration between various research groups participating to the COST Chemistry initiative of the EU

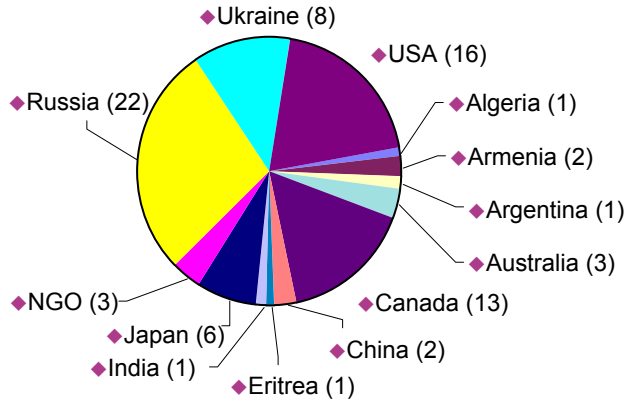
- In particular from the WGs of the Action D23: Metachem: Metalaboratories for Complex Computational Applications in Chemistry
- Also some national initiatives like GRID.IT by the Italian National Research Council have contributed on putting together expertise from various groups.

COST- Countries (2002) Participating Institutions

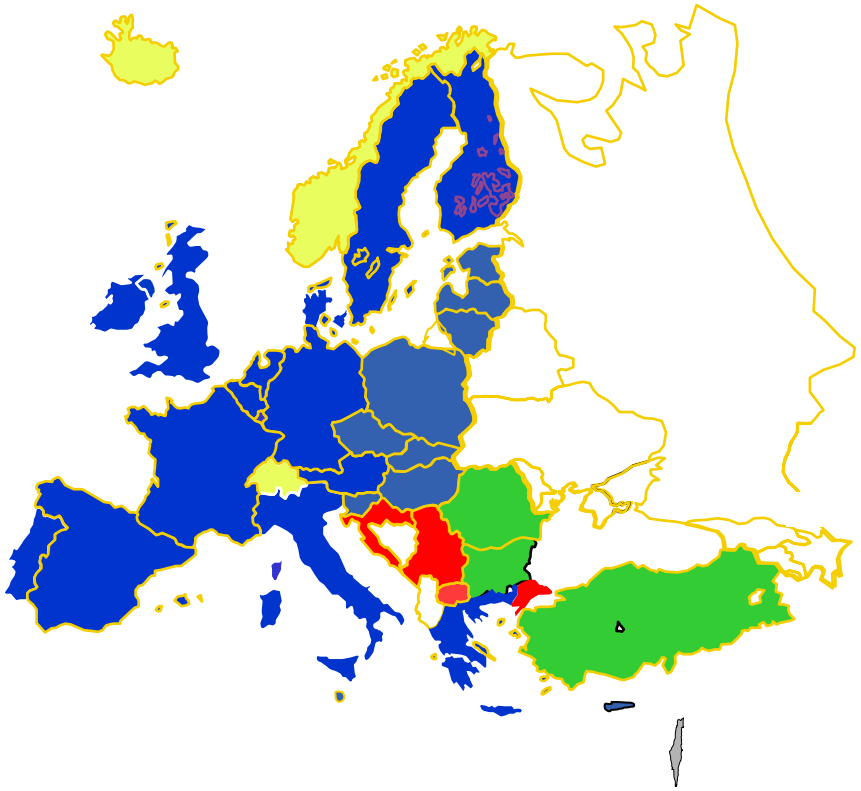




Participation of Non-COST Institutions (Total: 78)



COST Member States



- **The twentyfive EU Member States**
- **EFTA Member States**
  - Iceland
  - Norway
  - Switzerland*
- **Candidate Counties**
  - Bulgaria
  - Romania
  - Turkey *
- **Other Countries**
  - Federal Republic of Yugoslavia*
  - Former Yugoslav Republic of Macedonia*
  - Croatia *
- **Co-operating State**
  - Israel

* Not Associated to FP

The COST Chemistry domain

- Created in 1992
- 34 Actions launched since 1992 (12 completed with final evaluation)
- 22 running (10 had the mid term evaluation)
- Research Chemists from 31 COST Countries
- Associated Institutions from Australia, Japan, Russia, Ukraine and the USA

Metachem: Matalaboratories for Complex Applications in Chemistry

- **MURQM**: Multireference Quantum Chemical Methods
- **DIRAC**: Four Component Relativistic Quantum Chemical Calculations
- **SIMBEX**: Simulation of Molecular Beam Experiments
- **DYSTS**: Dynamics and Spectroscopy of Systems : Relevant to Environment and Applied Chemistry
- **ELCHEM**: E-learning Technologies for Chemistry
- **ICAB**: Integration of Codes for Ab Initio Methods

EU partners of D23 WGs

Simbex

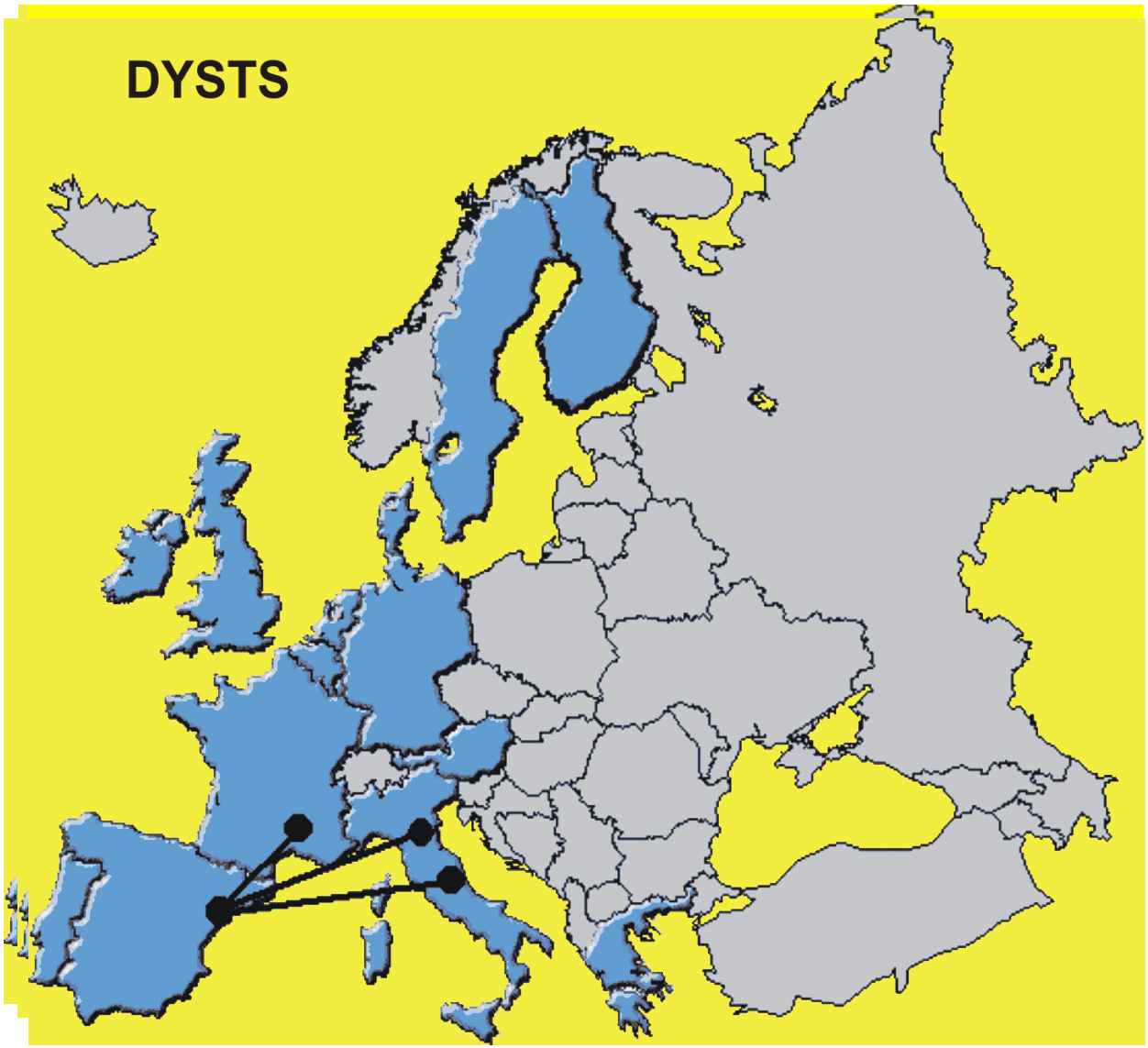
Murqm

Dirac

Elchem

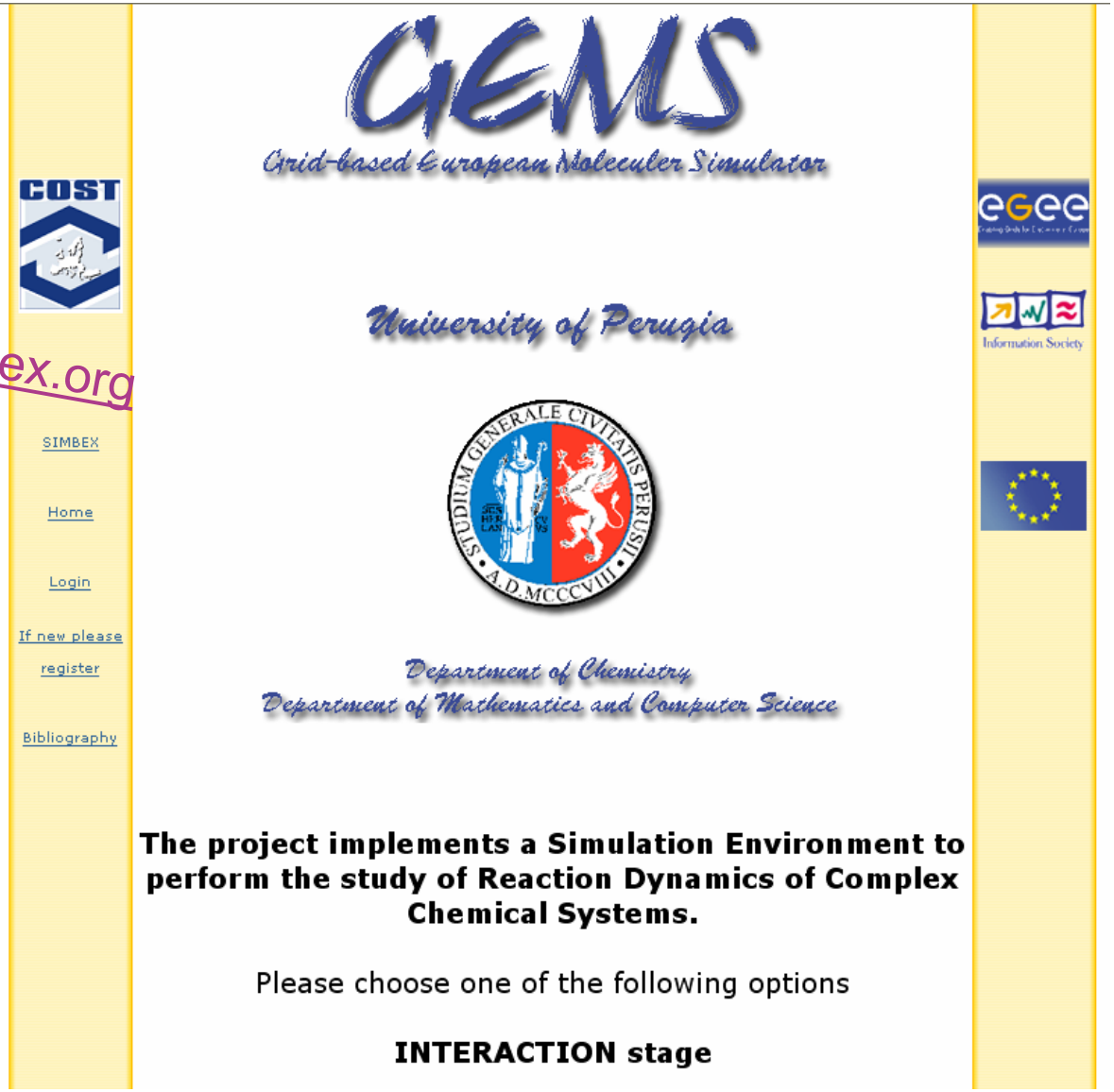
Icab

Dysts

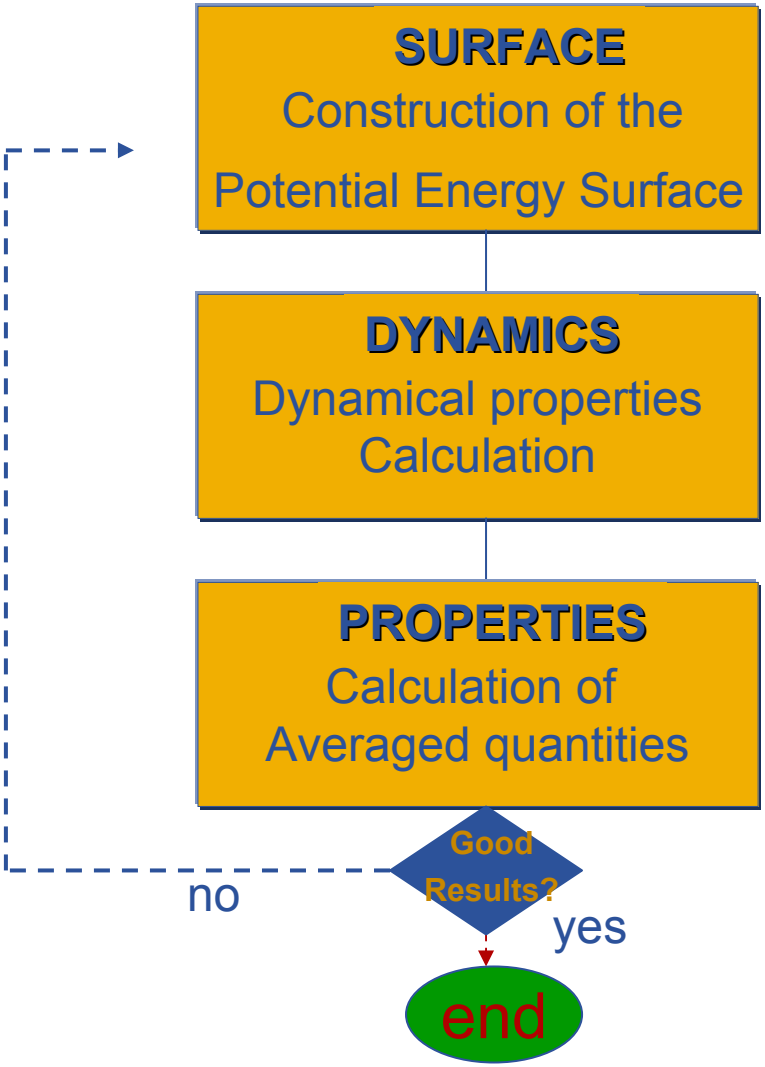


Grid based European Molecular Simulator (GEMS)

http://gems.simbex.org



The architecture of the Simulator



Electronic Structure calculation and representation

- **Isolated (gas phase) small molecules**
  - GAMESS-UK, GAMESS US, MOLPRO
- **Isolated (gas phase) large molecules**
  - (GAUSSIAN03)
- **Condensed phase and solid state calculations**
- **Large Molecules Topological analysis of interaction**
  - (AIMPAC, TOPOND)
- **Modeling and functional representations of the potential energy surfaces**
  - FITTING

Molecular Dynamics calculations

- **Exact quantum dynamics for small systems**
  - (TD)
- **Semiclassical and mixed classical-quantum for intermediate systems**
  - (ABCsem)
- **QM/MM and Car Parrinello**
  - (CPMD)
- **Classical dynamics**
  - (ABCtraj, VENUS; DL_POLY)

Observable Properties

- **Structure and stability calculations for aggregates of various sizes**
- **Kinetics and fluid dynamics calculations**
- **Thermodynamics properties**
- **Direct Monte Carlo calculations**
  - (DSMC)
- **Condensed phase and liquid crystals**
- **Cross sections and rate coefficients**
  - Energy and angular distributions

Critical Features of the Individual Programs

- AB INITIO METHODS (MOLPRO, GAMESS, ADC, GAUSSIAN,) resource requests are proportional to N^3 (N is the number of electrons) and to M^D (M is the number of grid points per dimension D) for CPU and disc demand.
- EMPIRICAL FORCE FIELDS (Venus, DL_POLY, …) resource requests are proportional to P! (P is the number of atoms)

Deployment of the VO

- **Even if all technical aspects have been implemented and we should able to start the deployment of the CompChem VO, we have chosen to begin in a stable environment. We have chosen so far to adopt GILDA certificates participating to the test bed infrastructure**
  - A **new Resource Centre** has been ported into GILDA (13 nodes + CE + SE)
  - A **User Interface** has been created to interface the user to the Grid
  - A **prototype** implementation of the Molecular Simulator has been deployed (two hours ago has been shown in the Demo session)
  - Also a prototype implementation of the Molecular Simulator has been deployed in **GENIUS**, thanks to the collaboration with INFN Catania and CNAF people (effective and wonderful experience!)

Issue related to ComChem VO

- **The implementation of the prototype has shown the most relevant aspects of GEMS**
  - **Interactive access** to the Grid through Sockets
    - We use the supported set of ports: 20000-25000, left **open by the firewall** in all nodes of the Grid
    - We need a **small latency** to schedule the job on the Grid
  - **Parallel** and distributed nature of most CompChem programs
    - **MPI prerequisite** in all nodes
    - We need the **highest** possible **number of Working Nodes**

Issue related to ComChem VO

- **Commercial Licence** of some programs
  - Some programs, especially those for the Ab Initio calculations (GAMESS, MOLPRO, MOLCAS, etc) require a Commercial Licence even if used for Research purposes.
  - A strategy for acquiring the licences for the users of the Grid may be sought.
  - Several sites own the licence for various programs in their machines. We need a mechanism to map the sites owning those programs, so that using this information the user may **specify such requisite in the JDL script**, when submitting the job to the Grid.

Populating the VO

- We will start right now the deployment of the VO, following the instruction received on Monday https://edms.cern.ch/document/503245
- A consortium agreement will be signed with us by the participating institutions.
- We will contribute to the Grid with new resources and expertise.
- A new COST D23 Action called GridChem will be started next year following the ending initiative MetaChem.

List of candidate members

Institution	#Researchers
Dept.Chemistry, University of Barcelona	5
Dept. Physical Chemistry, Univ. Basque Country	3
Institute of Chemistry, Hungarian Academy of Sciences	2
School of Chemistry, Bristol	3
Institute for Theoretical Chemistry and Structural Biology University of Vienna	3
Department of Chemistry, University of Bari,	3
Department of Chemistry, Univesity of Bologna	3
ISTM, Italian National Research Council, Perugia	6
Dept. of Chemistry, University of Perugia	5
COST working groups	27
TOTAL	**60**

Conclusions

- **Preliminary activities of the VO have been illustrated**
  - The experience inside GILDA VO
  - The deployment of a Resource Center
  - The prototype of GEMS
- **The main issues related to the implementation of the prototype of the simulator have been discussed**
  - Interactivity (via Sockets)
  - MPI as a requirement for CompChem jobs
  - The standard open ports of Grid firewalls are used
  - Commercial licences
    - Acquisition policy for the Grid users?
    - Propagation of related information described via JDL