



Enabling Grids for E-science

BioMed GridSchool 2009

11-17 May 2009, Varenna, Lake Como, Italy

BioMedGrid 2009 Summer School.

Luciano Milanese

*Institute of Biomedical Technologies CNR-ITB,
via F.lli Cervi 93, I-20090 Segrate (Milano), Italy
luciano.milanese@itb.cnr.it*

www.eu-egee.org



- The Biomed Grid School 2009 offers a hands on approach to Biomed applications in a grid computing environment.



- **Location**
- Villa Monastero, Via Polvani 2, Varenna, Italy
- **Registration**
- Register for the Biomed Grid School 2008 online:
<http://www.biomedgrid.it/registration>
- **Info**
- Further information is available at the following address:
- <http://www.biomedgrid.it>
- Contact the Biomed Grid school organisers at the following address:
- biomedgrid@itb.cnr.it

- Luciano Milanesi (Chairman), CNR-ITB, Italy
- Roberto Barbera, INFN, Univ. Catania, Italy
- Christophe Blanchet, CNRS, France
- Ignacio Blanquer, UPV, Spain
- Vincent Breton, CNRS, France
- David Fergusson, NeSC, UK
- Robin McConnell, NeSC, UK
- Johan Montagnat, CNRS, France
- Georgina Moulton, University of Manchester, UK

- Francesco Beltrame, dip. ICT, CNR, Italy
- Howard Bilofsky, University of Pennsylvania, USA
- Joerg Freund, Siemens, Germany
- Vicente Hernandez, UPV, Spain
- Martin Hofmann-APitius, University of Bonn, Germany
- Miklós Kozlovsky, Mta Sztaki, Hungary
- Mary Kratz, University of Michigan, USA
- Yannick Legre, Maat-G, France
- Charles Loomis, CNRS/LAL, France
- Giorgo Maggi, INFN, Univ. Bari, Italy
- David Manset, Maat-G, Spain
- Mirco Mazzucato, INFN, Italy
- Richard Mc Clatchey, University of West England, UK
- Joel Saltz, Ohio State University, USA
- Jonathan Silverstein, University of Chicago, USA
- Gergely Sipos, Mta Sztaki, Hungary
- Tony Solomonides, University of West England, UK
- Domenico Talia, University of Calabria, Italy
- Manolis Tsiknakis, Forth, Greece

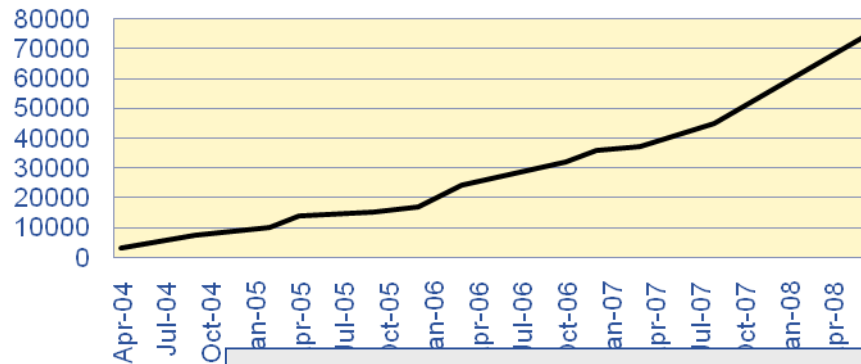


- **The Biomed Grid School will give bioinformatics and biomedical developers the opportunity:**
 - to learn to use new services and functionalities made available by new middleware releases (gLite)
 - to implement biomedical applications in the EGEE – GRID environments.
 - to build high level interfaces for GRID and Web Services for data provision and management.
 - to improve skills in Web Services, Workflow technology and GRID services for Bioinformatics e Biomedical applications.
 - to learn how to generate complex workflows for Bioinformatics and Medical Imaging analysis
 - to discuss with the experts of the life science related projects

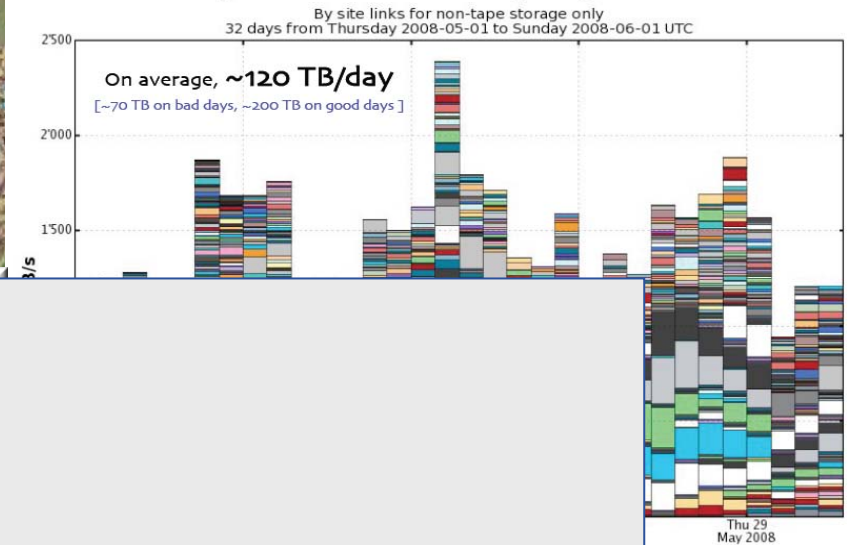
- **Basic concepts of Grid**
 - gLite services
 - Job execution
 - Workflow development
 - Parameter study development on EGEE
- **Implementation of the Biomedical applications**
 - Portals
 - User Interface
 - Monitoring tools
 - Data Base end Metadata
- **Further information**
 - How to get access to EGEE
 - How to port your own application to EGEE



No. Cores



Daily CMS PHEX transfer rate, Debug + Production



- > 200 sites in 40 countries
- ~ 38 000 CPUs
- ~ 5 PB storage
- 98k jobs/day
- > 200 Virtual Organizations
- ↳ The world's largest multi-disciplinary Grid

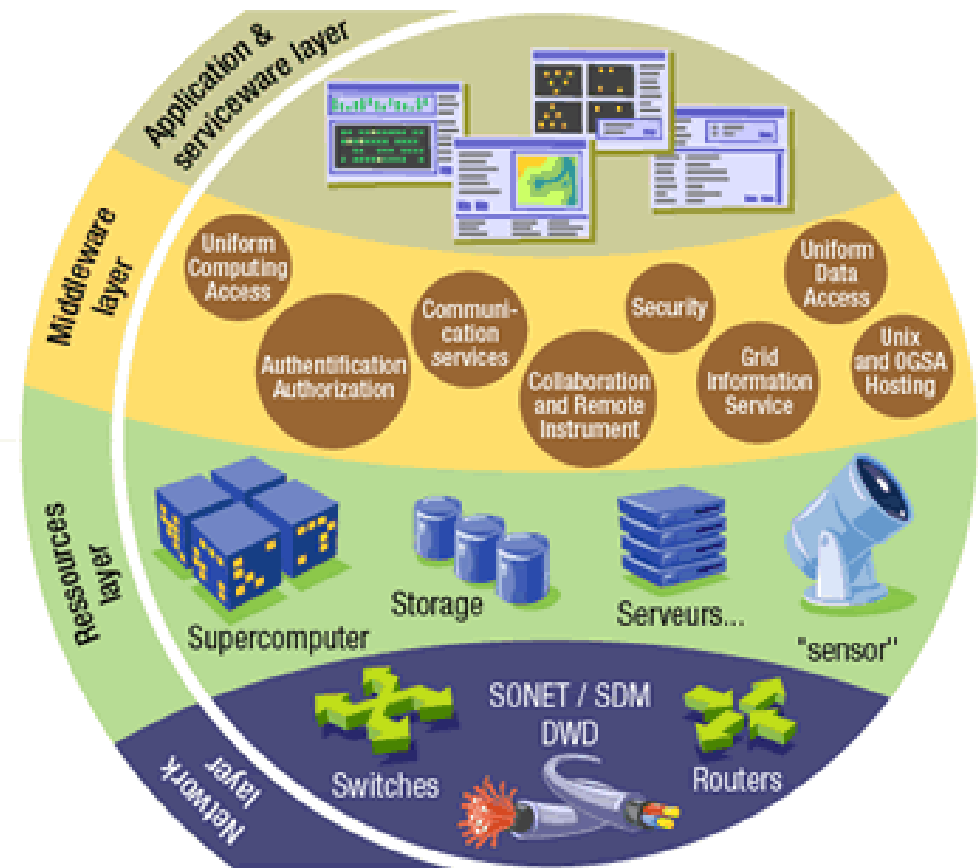
300
250
200
150
100
50
0

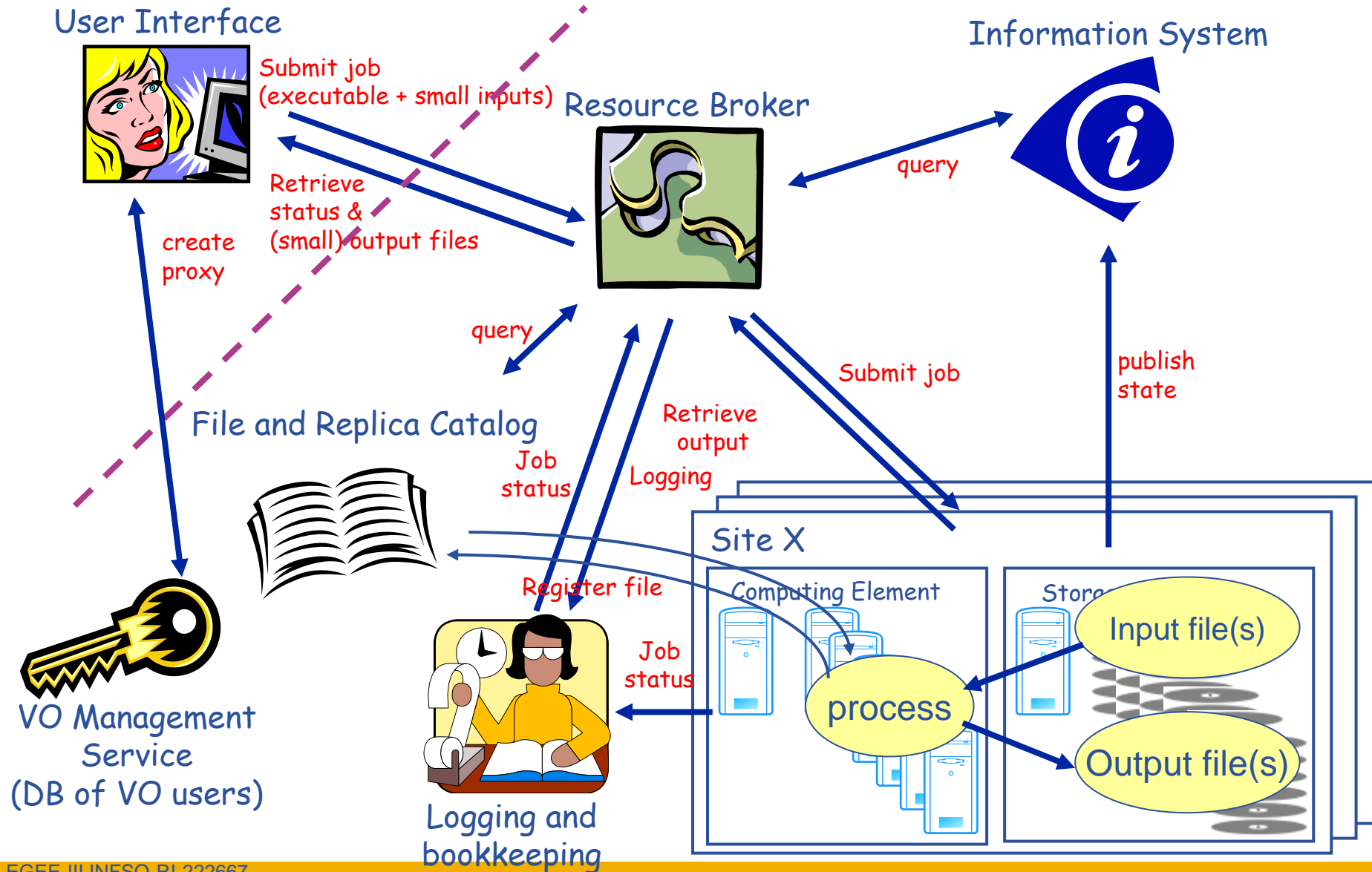
Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr

- The Grid relies on advanced software, called **middleware**, which interfaces between resources and the applications

- **The Grid middleware:**

- Basic services
 - Secure and effective access to resources
- High level services
 - Optimal use of resources
 - Authentication to the different sites that are used
 - Job execution & monitoring of progress
 - Problem recovery
 - Transfer of results back to the user







User Interface (UI): The place where users logon to the Grid



Resource Broker (RB) (Workload Management System (WMS)):
Matches the user requirements with the available resources on the Grid



Information System: Characteristics and status of CE and SE



File and replica catalog: Location of grid files and grid file replicas



Logging and Bookkeeping (LB): Log information of jobs

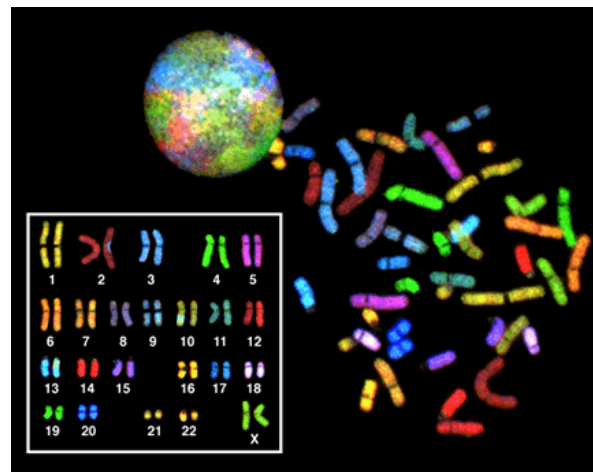


Computing Element (CE): A batch queue on a site's computers where the user's job is executed

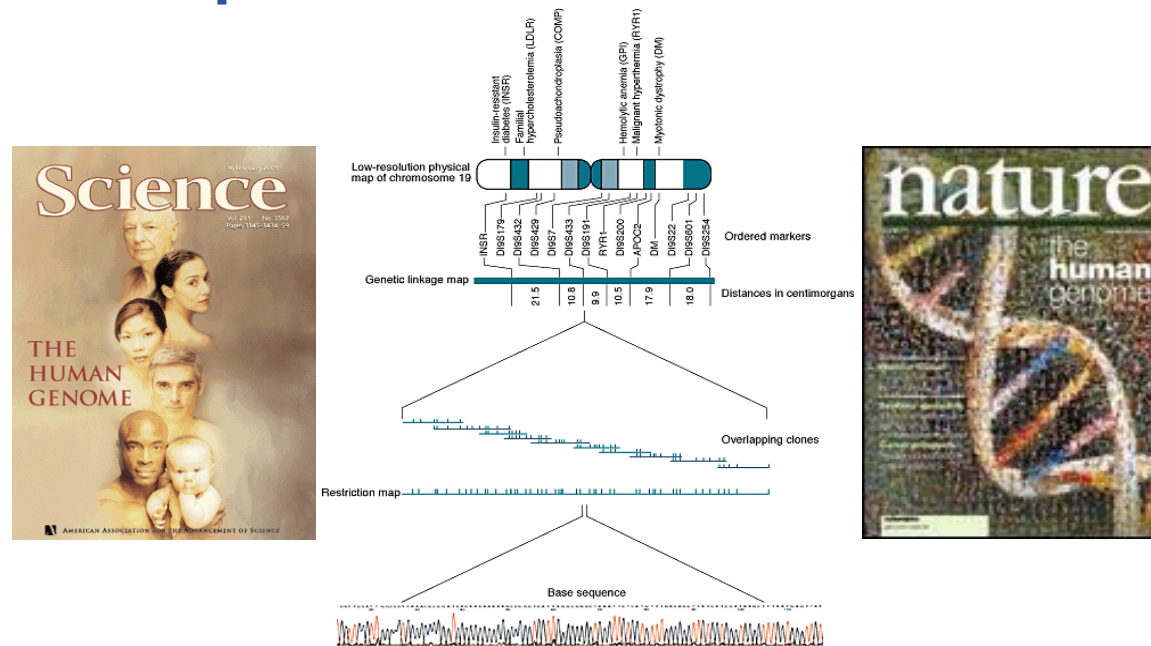


Storage Element (SE): provides (large-scale) storage for files

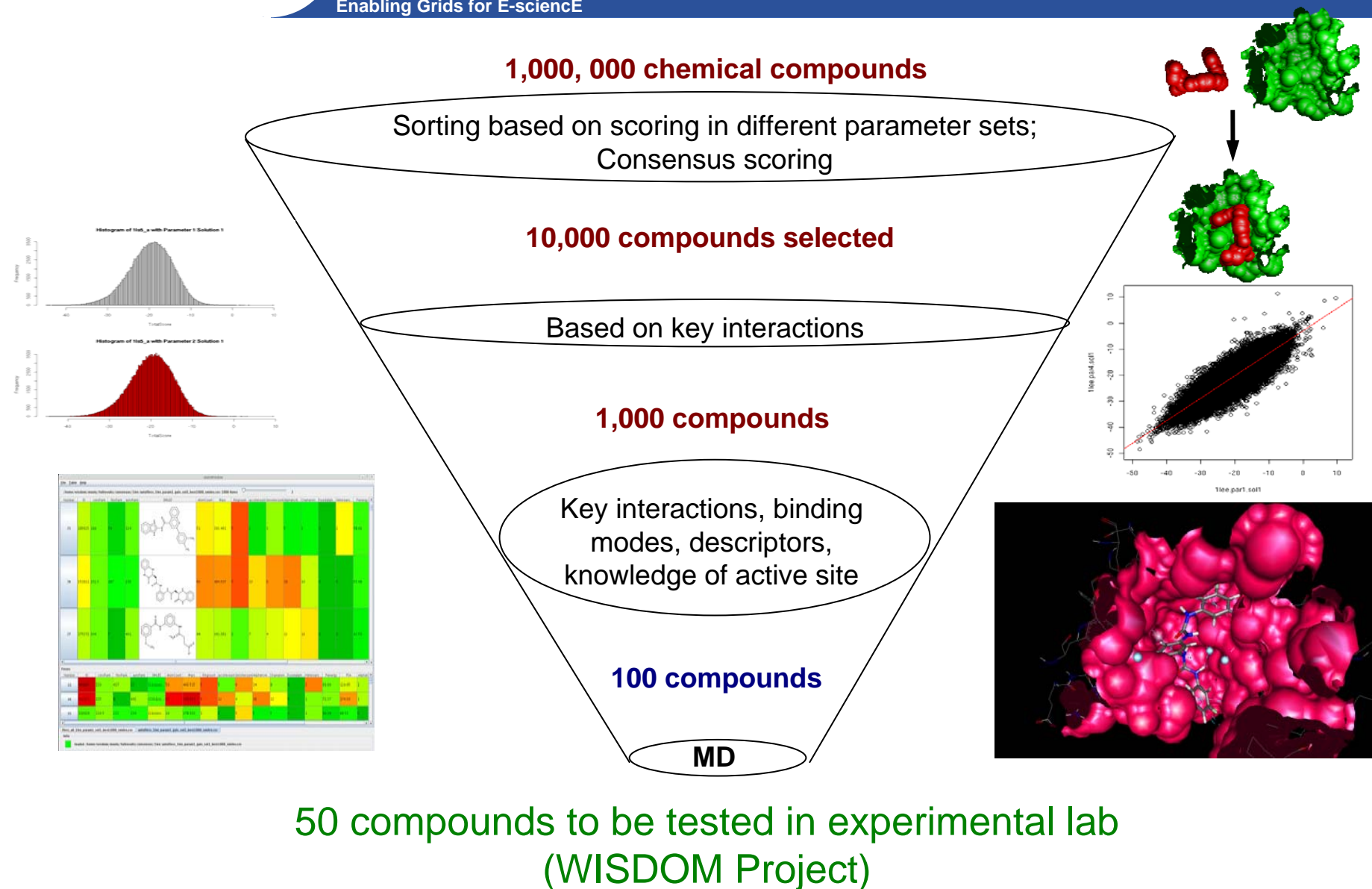
- The use of the new biological and biomedical technological platforms **need to deal with the increasing amount, complexity, and heterogeneity of biological and biomedical data.**
- Bioinformatics and Biomedical applications **have become an ideal research area where scientists can apply and further develop new intelligent computation methods, in both experimental and theoretical cases.**

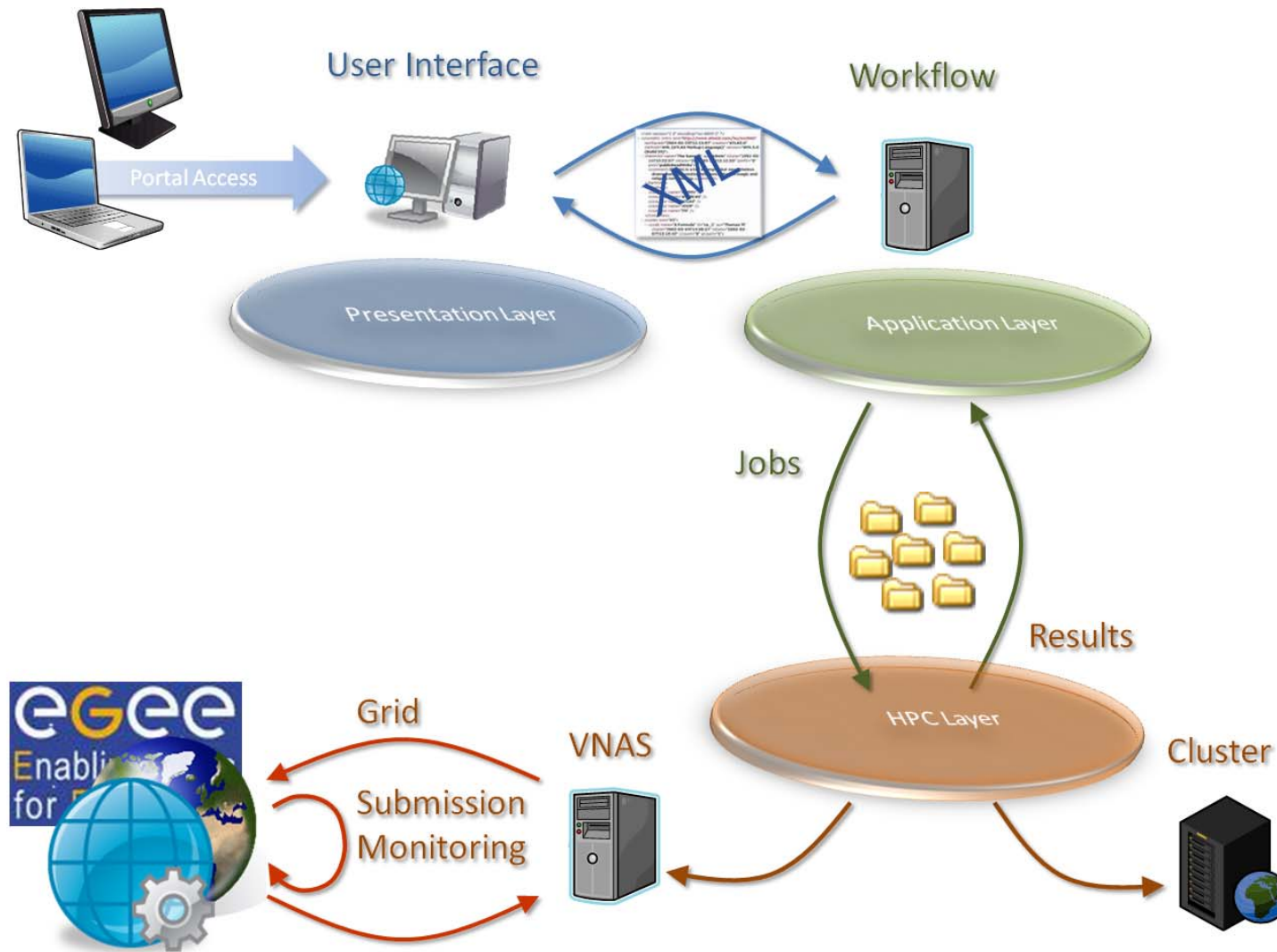


- “Post-genomics” focuses on the new tools and new methodologies emerging from the knowledge of genome sequences.
- Production and use of DNA micro arrays, analysis of transcriptome, proteome, metabolome are the different topics developed in this class.



- **A key development in the computational world has been the arrival of de novo design algorithms that use all available spatial information to be found within the target to design novel drugs.**
- **Coupling these algorithms to the rapidly growing body of information from structural genomics together with the new ICT technology (eg. HPC, GRID, Web Services, ecc.)**
- **provides a powerful new possibility for exploring design to a broad spectrum of genomics targets, including more challenging techniques such as:**
- **Gene discovery, SNP and Mutation, Protein interactions, Drug Discovery, Molecular Dynamics, System Biology, Gene Network ecc.**





Enabling Grids for E-science



Download Center

The comprehensive Universal Protein Resource download center.

(Current release note)

UniProt Download Center

Data Format: XML, FASTA, FLAT FILE

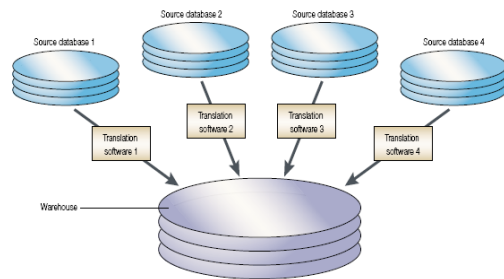
Search: Enter Text Here [Go] [Reset] [Advanced Search]

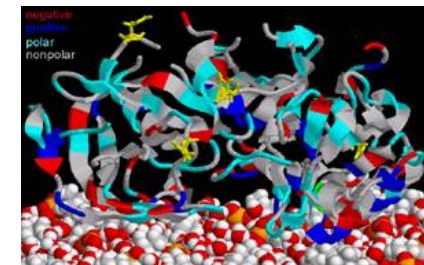
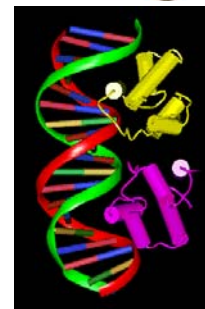
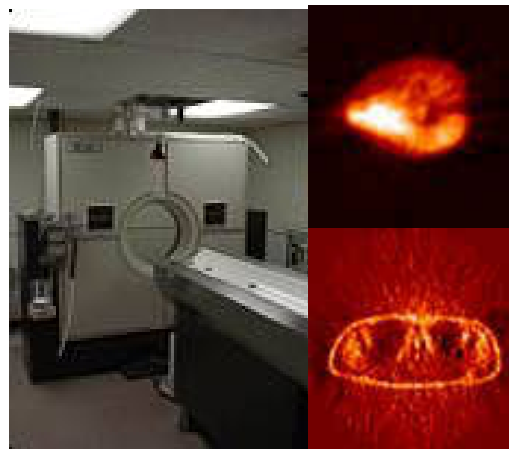
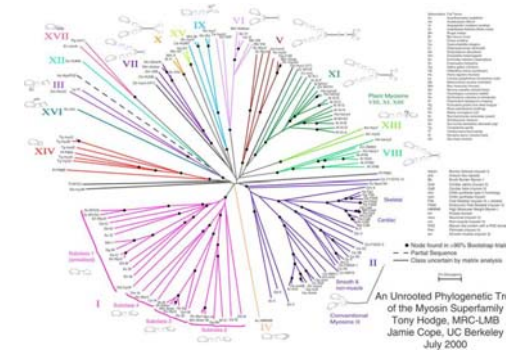
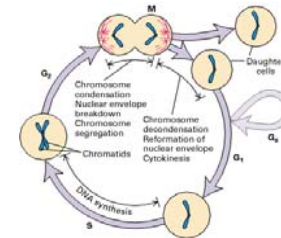
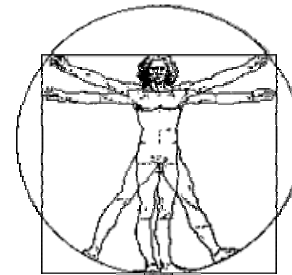
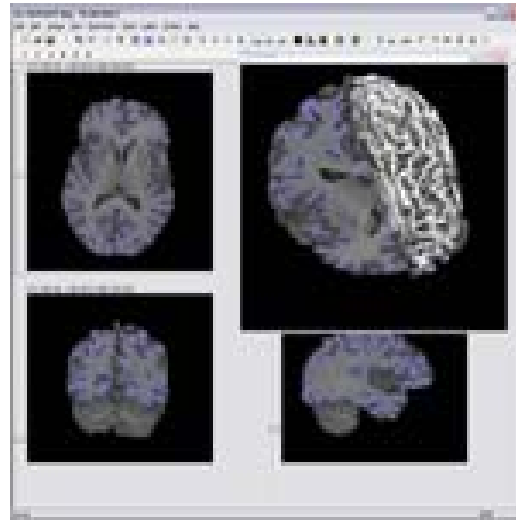
Databases

- InterPro
- UniProt
- PROSITE
- Pfam
- PRINTS
- ProDom

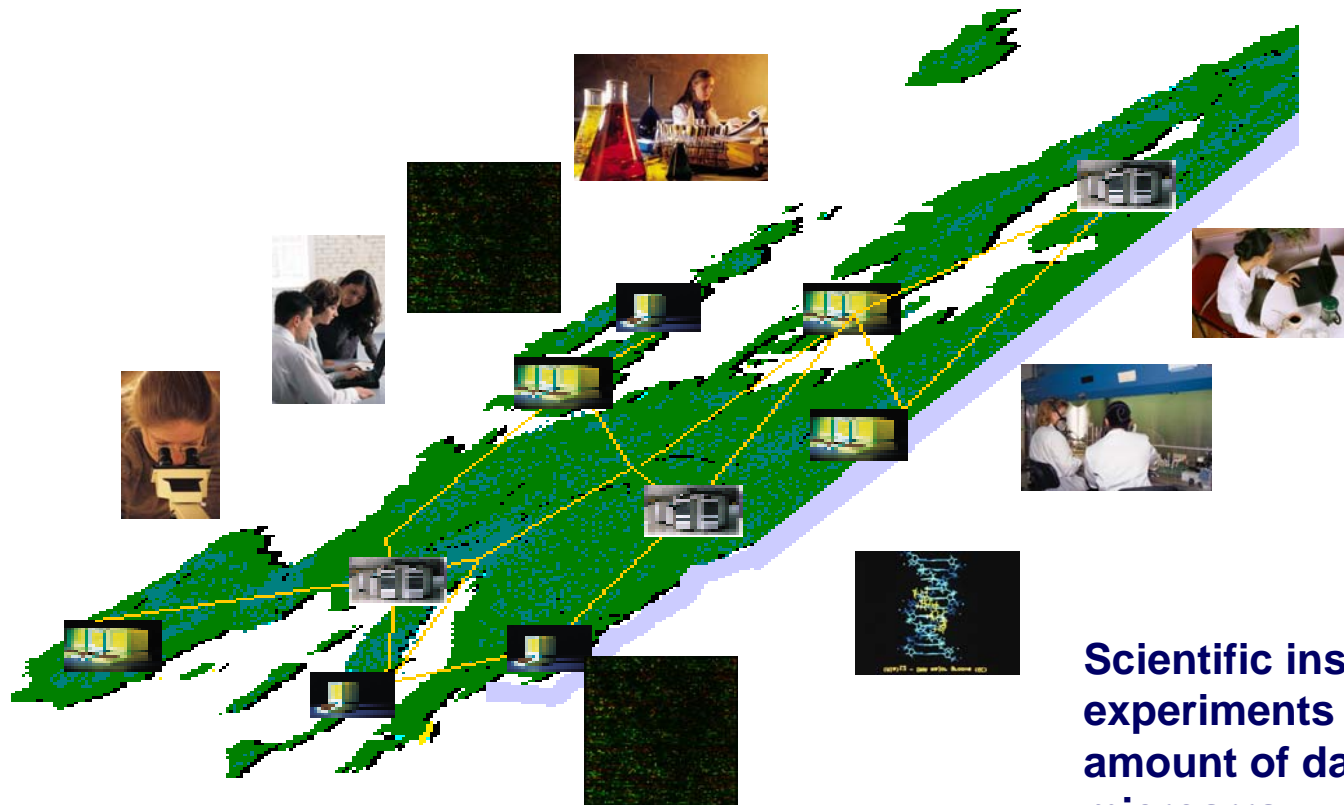
Major resources available by ftp (ftp.ncbi.nih.gov):

- BLAST Basic Local Alignment Search Tool
- CnSD
- Data Repository
- GenBank
- Gene
- Genome Assembly/Annotation Projects





Data analysis specific for bioinformatics allow the user to store and search genetics data, with direct access to the data files on distributed servers.



Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

Scientific instruments and experiments provide huge amount of data from microarray

- **EGEE**
 - <http://www.eu-egee.org/>
- **gLite middleware**
 - <http://www.glite.org>
- **gLite manuals, documentation**
 - <http://glite.web.cern.ch/glite/documentation/>
(gLite user guide)
- **Recommended External Software Packages for Egee CommuniTies (RESPECT)**
 - <http://egeena4.lal.in2p3.fr/>
- **Description of work of EGEE-III**
 - <https://edms.cern.ch/document/886385/4>

- The BioMedGrid Summer School 2009 has been supported by the EGEE III and related projects.



- And by the CNR-BIOINFORMATICS and by the LITBIO, ITALBIONET FIRB-MIUR projects.

