Vinod Kasam

How Grids can help in enabling complex computationally intensive biomedical applications

22/01/2007

Forschungszentrum Juelich

Bridge

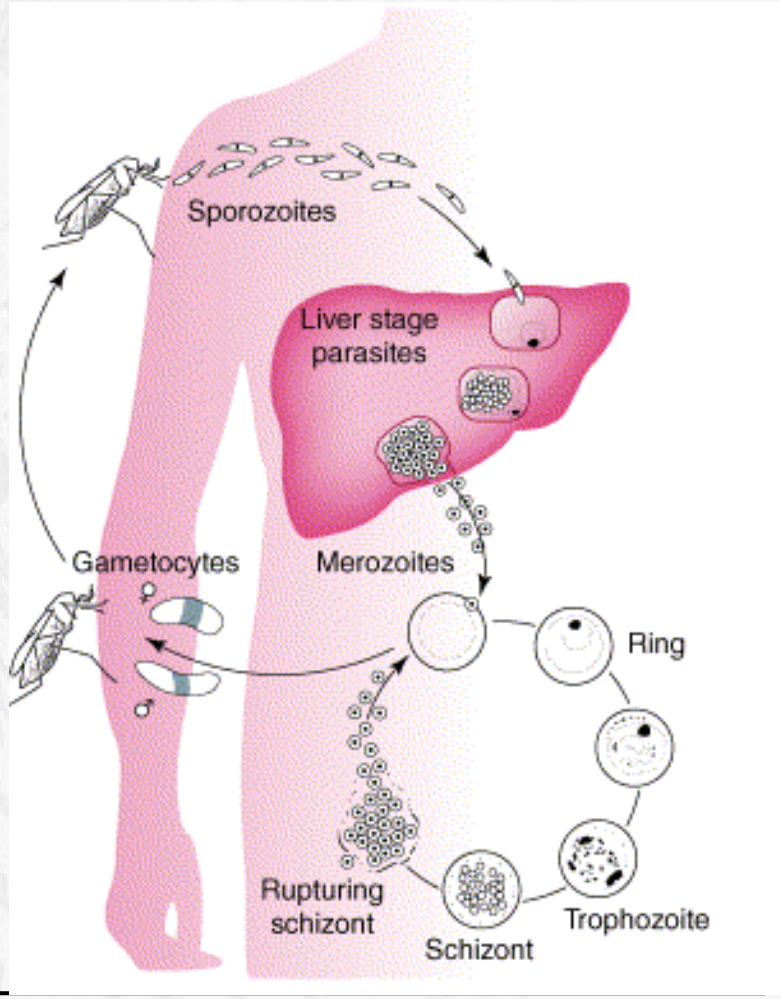
Information Society Technologies

Outline

- Grid enabled complex workflows
  - Part I : WISDOM-I
  - Part II: WISDOM-II
  - Part III: Molecular dynamics on Grid
- Future works

Motivation

- ~300 millions people affected in the world
- ~1 million people die every year
- A child die every 30 seconds
- Caused by protozoan parasites of the genus *Plasmodium*



Bridge

Information Society Technologies

FlexX and Amber

- FlexX: ~1-2 minutes per docking
- One executable: Can be Distributed
- Embarrassingly parallel

- Amber: Depending upon the simulation parameters
- ~15-20 minutes (WISDOM)
  - ~30-60 minutes
  - ~few hours

- Can be Embarrassingly parallel or massively parallel

- Interdependent several executables: Can be Parallelized

Wisdom I workflow

Grid

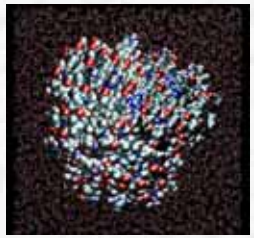
FlexX

1 million

Molecular docking

Molecular dynamics

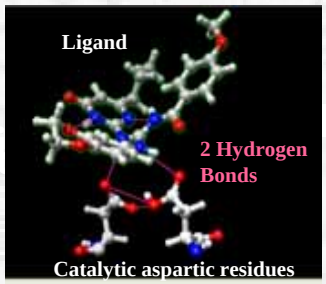
Amber



5000

Re-ranking
MMPBSA-GBSA

Chimera



180

Complex visualization

Wet Laboratory

30

in vitro

10

in vivo

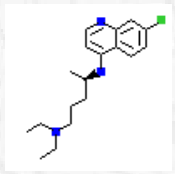
Bridge

Grid Enabled Virtual Docking in WISDOM-II

Millions of chemical compounds available

High Throughput Screening
1-10$/compound. Very expensive

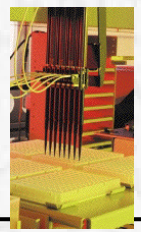
Chemical compounds
(ZINC databse): 4.3 million
Chembridge ~300,000



Targets (PDB): PvDHFR
PfDHFR, GST, tubulin

Molecular docking (FlexX)
~413 CPU years, 1.738 TB data
~100,000 dockings per one hour

Data challenge on EGEE
~90 days on ~5000 computers



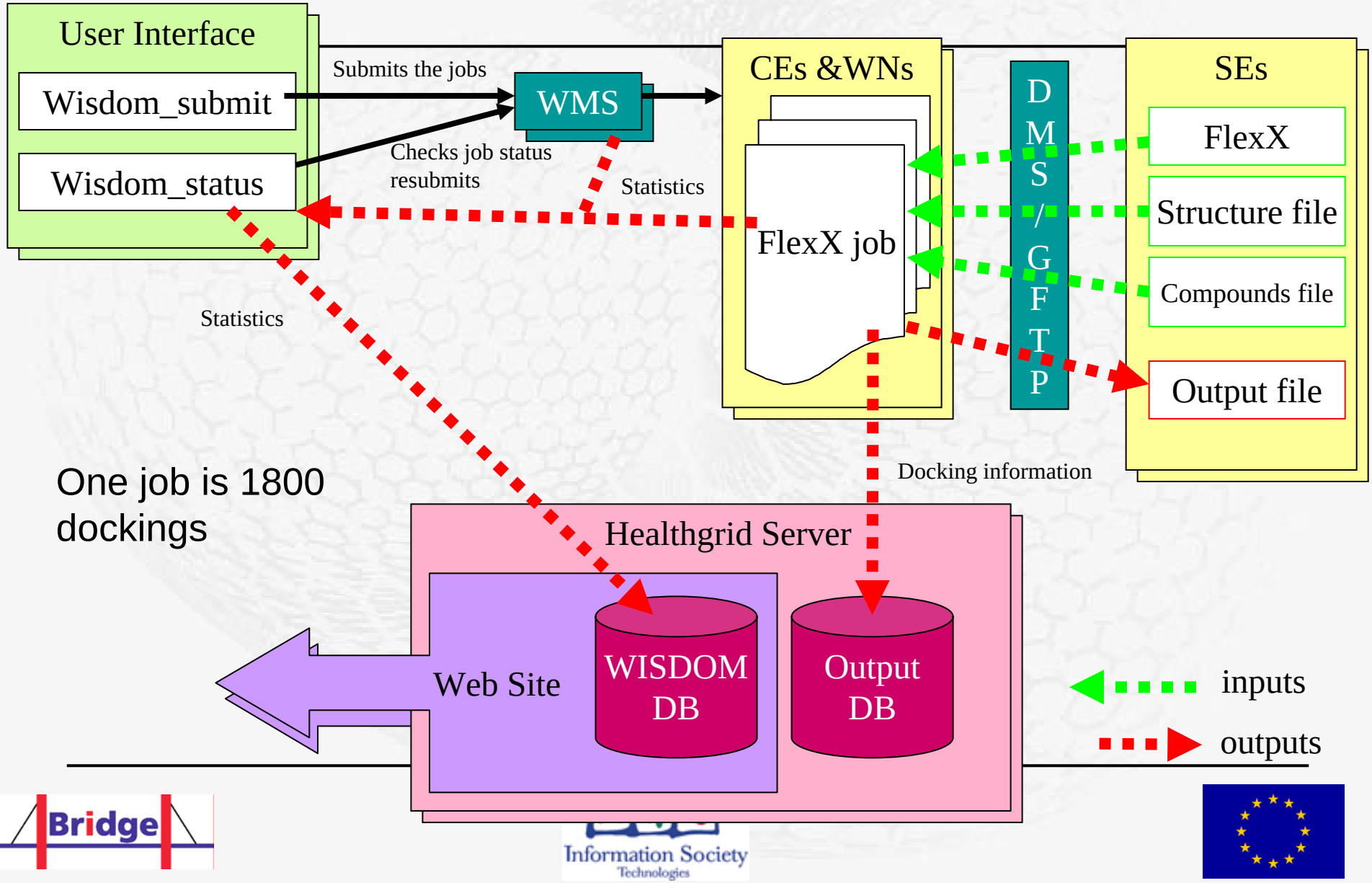
Hits screening using assays performed on living cells

- Leads
- Clinical testing
- Drug

Bridge

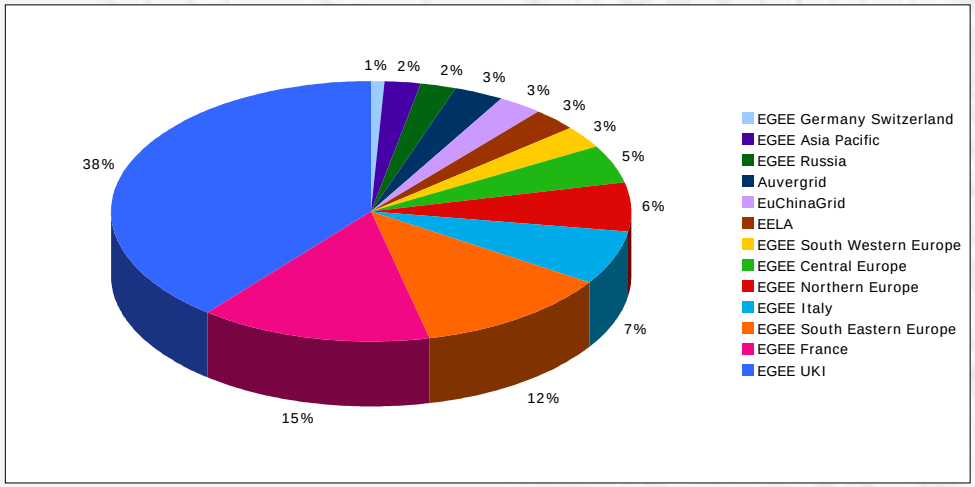
Information Society Technologies

Production environment



WISDOM-II: Grid statistics

Target Structures	Number of instances deployed
GST (A chain)	4 on EGEE
GST (B chain)	4 on EGEE
2BL9 (P. vivax wild type DHFR)	3 on EGEE, 1 on EELA
2BLC (P. vivax double mutant DHFR)	3 on EGEE, 1 on Auvergrid
Dm_vivax (P. vivax DHFR 2BLC minimized)	4 on EGEE
Wt_vivax (P. vivax DHFR 2BL9 minimized)	4 on EGEE
1J3K (P. falciparum Quadruple mutant DHFR)	4 on EGEE
1J3I (P. falciparum Wild type DHFR)	3 on EGEE, 1 on EuChinaGrid



Number of Jobs	77,504
Total Number of completed dockings	156,407,400
Estimated duration on 1 CPU	413 years
Duration of the experiment	76 days
Average throughput	78,400 dockings/hour
Maximum number of loaded licences (concurrent running jobs)	5,000
Number of used computing elements	98
Average duration of a job	41 hours
Average crunching factor	1,986
Volume of output results	1,738 TB

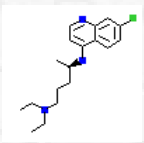
Bridge

Information Society Technologies

Grid enabled Molecular dynamics refinement

Compared to docking, Molecular dynamics needs more Computing resources

25000 best compounds
from docking
based on scoring and
docking pose



Target (PDB):
Plasmepsin II 1lee, DHFR, GST

Molecular dynamics (Amber 9)
~347 CPU years

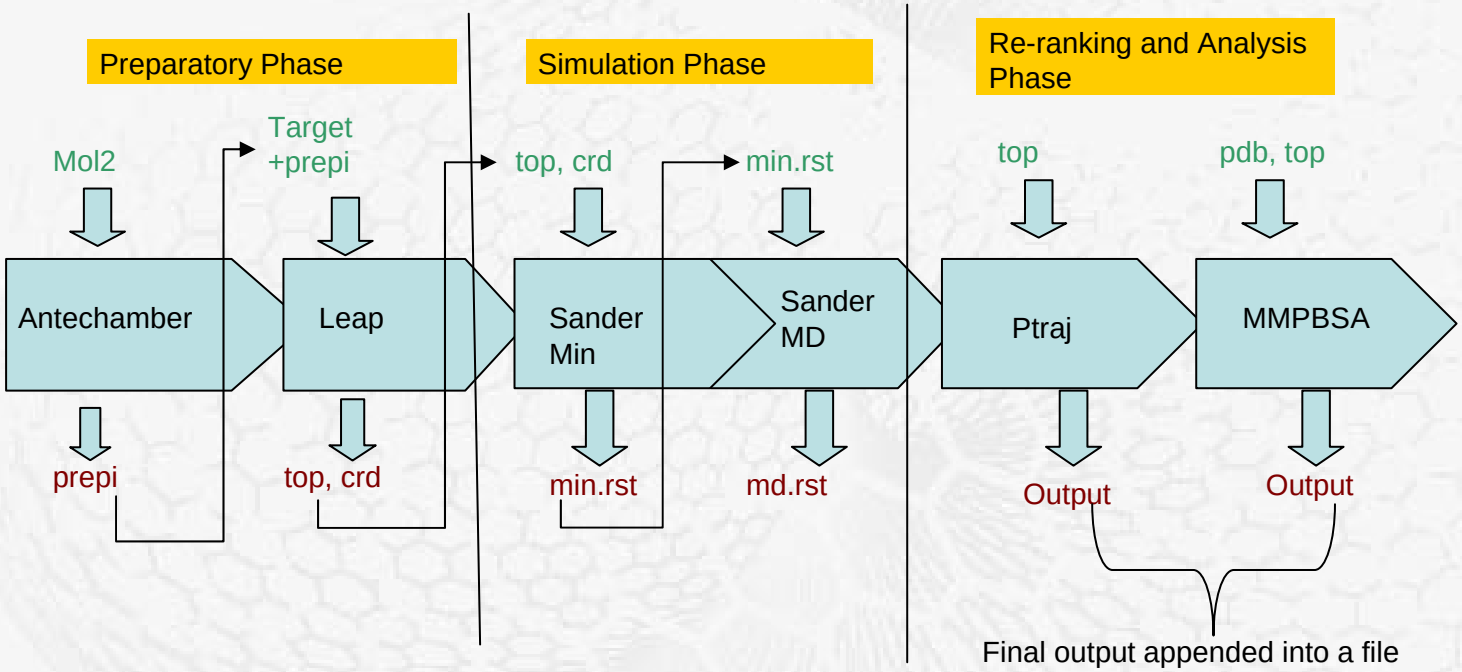
Deployment on Auvergrid
~25 days on 600 CPU

In vitro testing:
of best 30 compounds
(plasmepsin)

A. Ferrari, G. Degliesposti, M. Sgobba, G. Rastelli. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. Bioorganic & Medicinal Chemistry. 2007.

Refinement procedure developed by Giulio Rastelli's Team, Modena Univ, Italy

Amber software architecture



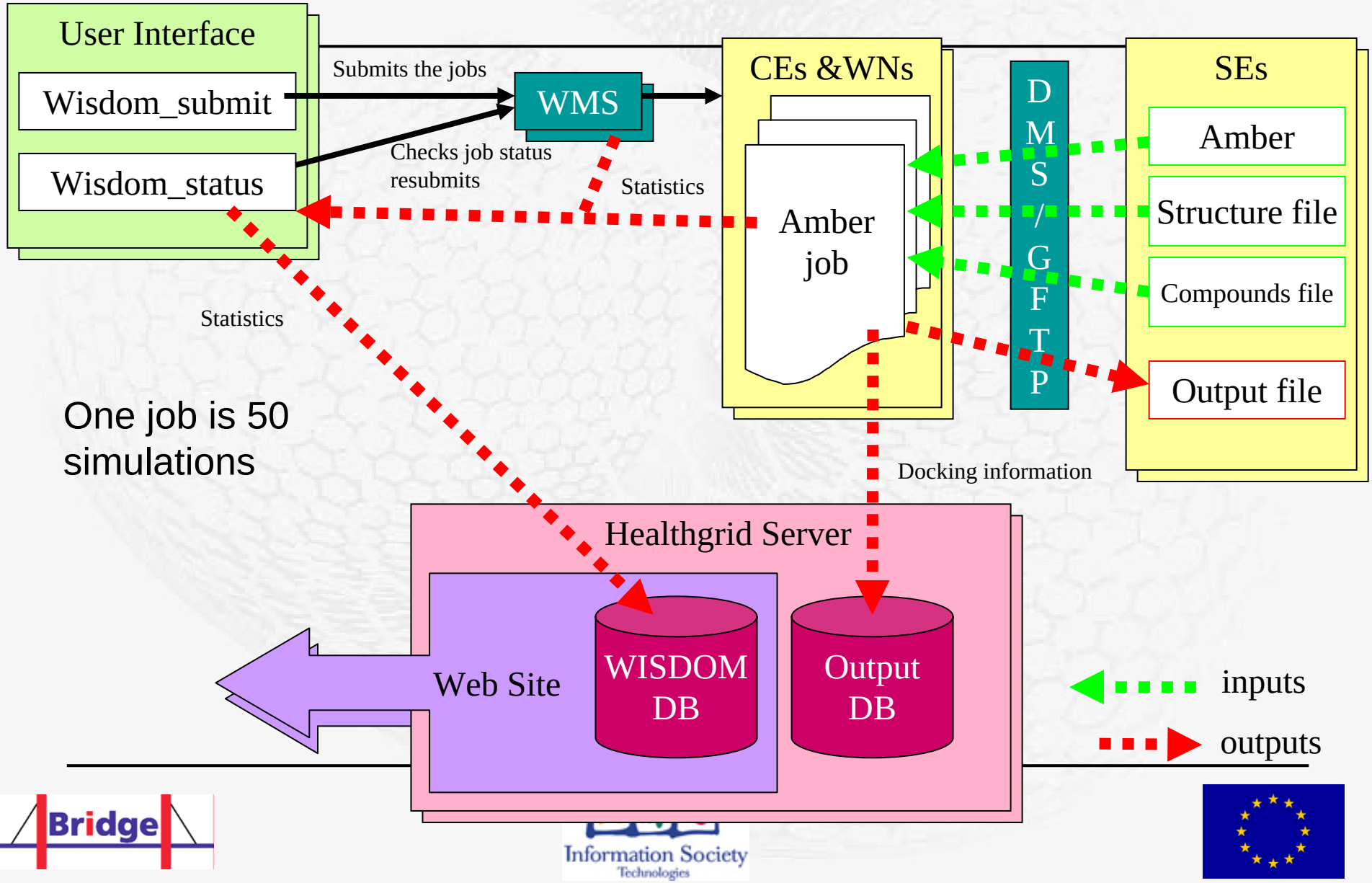
Software architechture

Management of the Interdependent steps is challenging

Deployment strategy

- Deployment is segmented into 3 phases
  - Storage Element of the grid
    - 5000 compounds: 100 jobs
    - Bash Script to the run the Amber job
  - Run the jobs
    - Running the jobs on Clermont cluster by using WISDOM production environment
  - Checking the status of the jobs
    - Done automatically (developed for WISDOM data challenges)

Production environment



Licensing and issues

- Licensing
  - Flexlm server is used: For FlexX
  - License is integrated: For Amber9
- Issues
  - Use of Amber restricted to grid users
  - No licensing issues for FlexX
- Issue addressed
  - Access granted to all the nodes where Amber is installed

Future works

- In context to Pharma center
  - Establish “Grid enabled complex work flows” at university of Bonn
  - Biological side: collaboration with pharmaceutical department,University of Bonn
  - Grid Resources: “OPEN for collaboration”

WISDOM project

- Two huge CPU demanding applications were deployed.
- Identified and tested few in experimental laboratories
- *Proof of concept: success achieved against plasmepsin*

Would have not been possible without computational grids

Bridge

Information Society Technologies

Bioinformatian interest in Grid Computing

- Grids are unique tools for
  - **Collecting and sharing information**
  - **Mobilizing resources routinely or in emergency**
  - **Perform CPU intensive tasks**

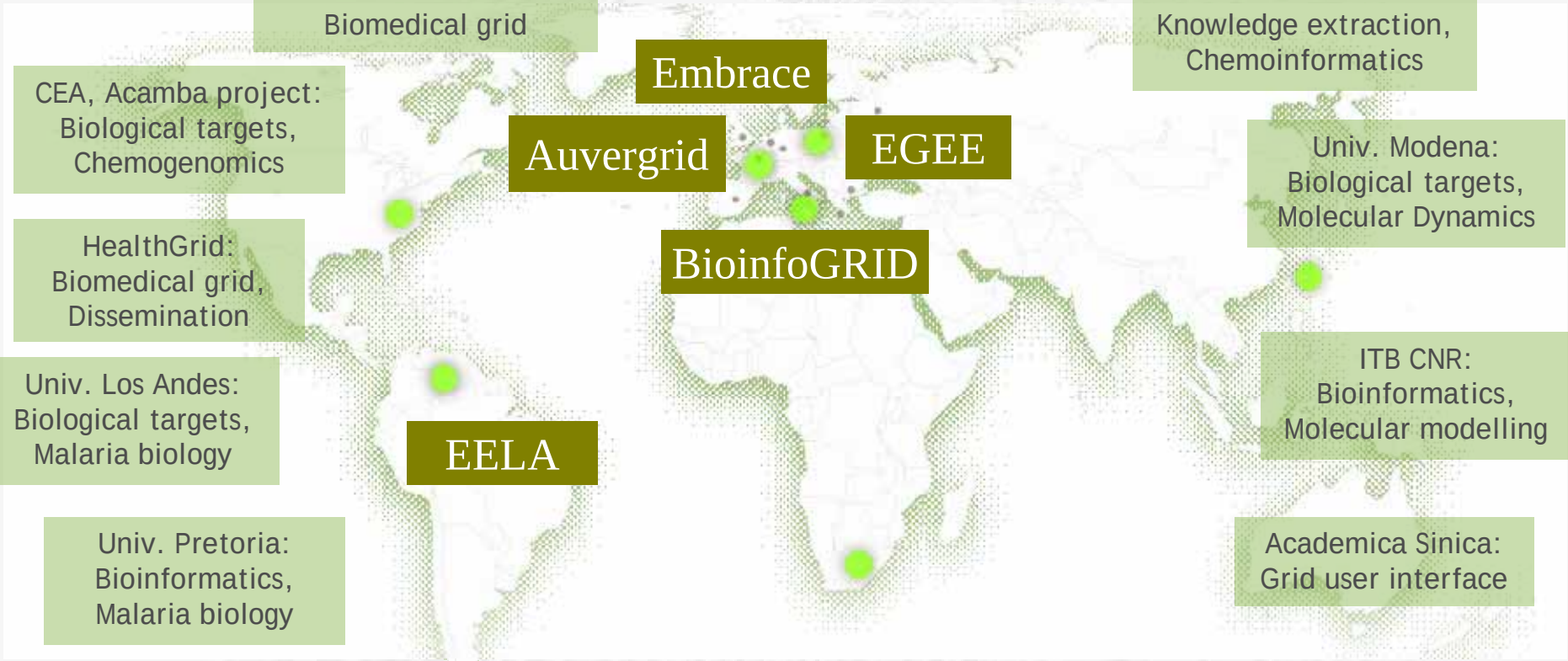
Inter country and Inter continental resource sharing is done during WISDOM project

Grid added value in disease management

- Deployment of infrastructures (federation of databases) to collect biomedical data and improve disease monitoring
- Cross-organizational collaboration space to share data and resources
- Challenges
  - Infrastructure capacity building in developing and poor countries
  - Grid technology must provide the services for data and knowledge management
  - IT expertise and willingness to share information is needed from the participating healthcare centers

Credit: V. Breton

Long term vision: a grid for emerging diseases



Use the grid technology to foster research and development on malaria and other neglected diseases

Contacts also established with WHO, Microsoft, TATRC, Argonne, SDSC, SERONO, NOVARTIS, Sanofi-Aventis, Hospitals in subsaharian Africa,

Bridge

Information Society Technologies

Acknowledgement

- Martin Hofmann-Apitius, Marc Zimmermann a
- Vincent Breton, Nicolas Jacq, Jean Salzeman (France)
- Giulio Rastelli and his group (Italy)
- Doman Kim and his group (South Korea)
- Wisdom Collaboration

Any Questions??

Bridge

Information Society Technologies