DigitalWorld
Cancun 1 - 7 February 2009

Distributed
European
Infrastructure for
Supercomputing
Applications

The Evolution of e-Infrastructures in Research, Industry, and Education

Wolfgang Gentzsch
The DEISA Project & Board of Directors of OGF
gentzsch at rzg.mpg.de

"It's hard to make predictions, especially about the future"

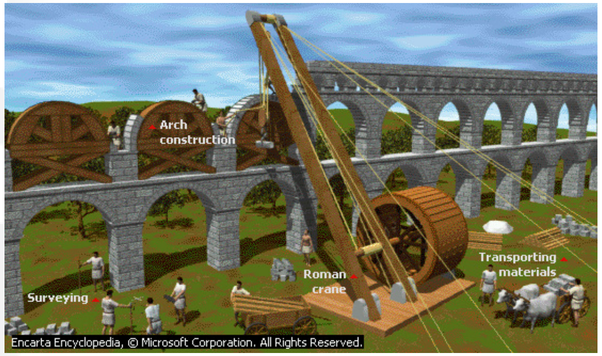
Yogi Berra

Content

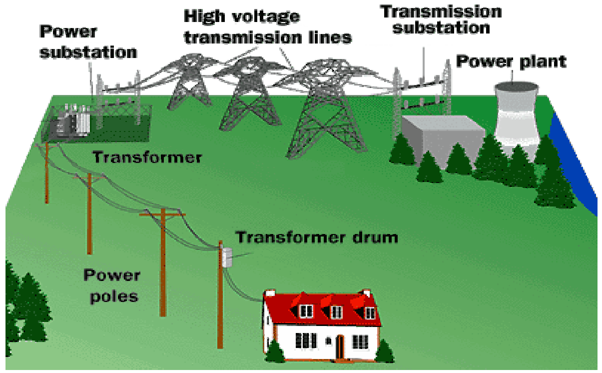
- Service Infrastructures: nothing new
- Building Blocks: Computers, Grids, Clouds
- Example: DEISA Ecosystem for Science Applications
- Next-Generation e-Infrastructures
  - e-Learning on e-Infrastructures = e-Learning 2.0

Distributed
European
Infrastructure for
Supercomputing
Applications

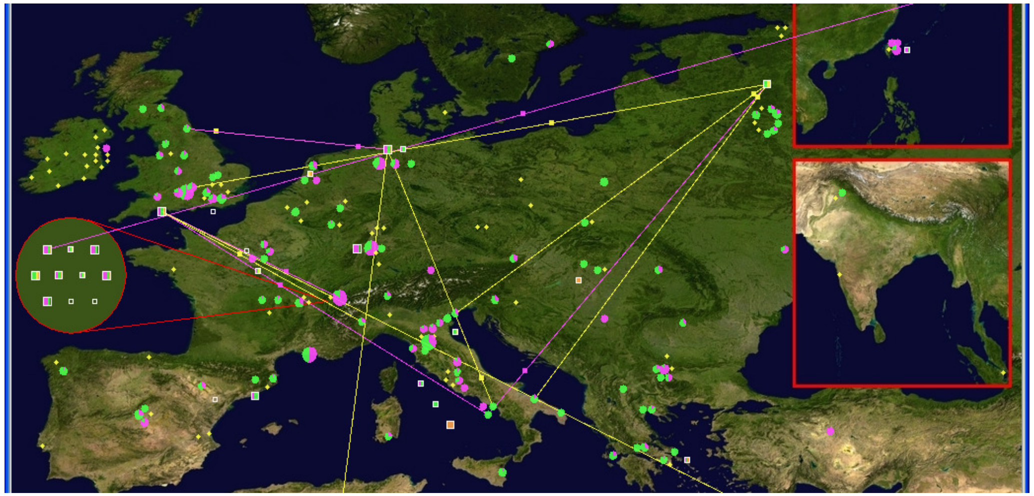
Service Infrastructures



Ancient Rome: ten aqueducts, some 150,000 m^3 of water each day



Electrical Power Grid Infrastructure



EGEE – Enabling Grids for E-SciencE

HPC Centers

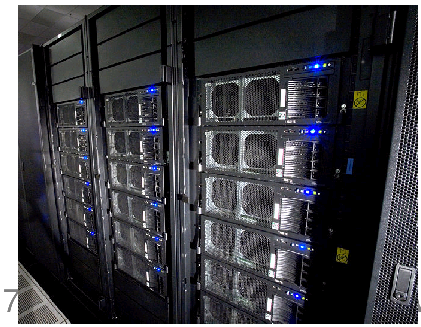
Distributed European Infrastructure for Supercomputing Applications

- HPC Centers are **service providers**, for past 35 years
- Computing, storage, applications, data, etc IT services
- Serve (local) research, education, and industry
- Very professional: to end-users, they appear almost as Cloud services (AWS Cloud definition: easy, secure, flexible, on demand, pay per use, self serve)
- But: no virtualization, semi-automatic, static
- They could become a Cloud customer for dynamic scaling and adopting to changing business and user demands

Distributed
European
Infrastructure for
Supercomputing
Applications

RoadRunner, Today's Fasted SC

- **1986**, Cray-2 breaking the **Giga**flop/s barrier
- **1997**, Intel ASCI Red, breaking **Tera**flop/s barrier
- **2008**, IBM RoadRunner, breaking **Peta**flop/s
  - At DOE's Los Alamos National Laboratory
  - 1.026 Linpack Petaflop/s solving 2 Mio equations
  - 6912 dual-core Opteron & 12960 IBM Cell
  - #1 on the Top500 list of June 2008



Grids

Distributed
European
Infrastructure for
Supercomputing
Applications

1998: The Grid: Blueprint for a New Computing Infrastructure:

"... hardware and software infrastructure ... dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities."

2002: The Anatomy of the Grid:

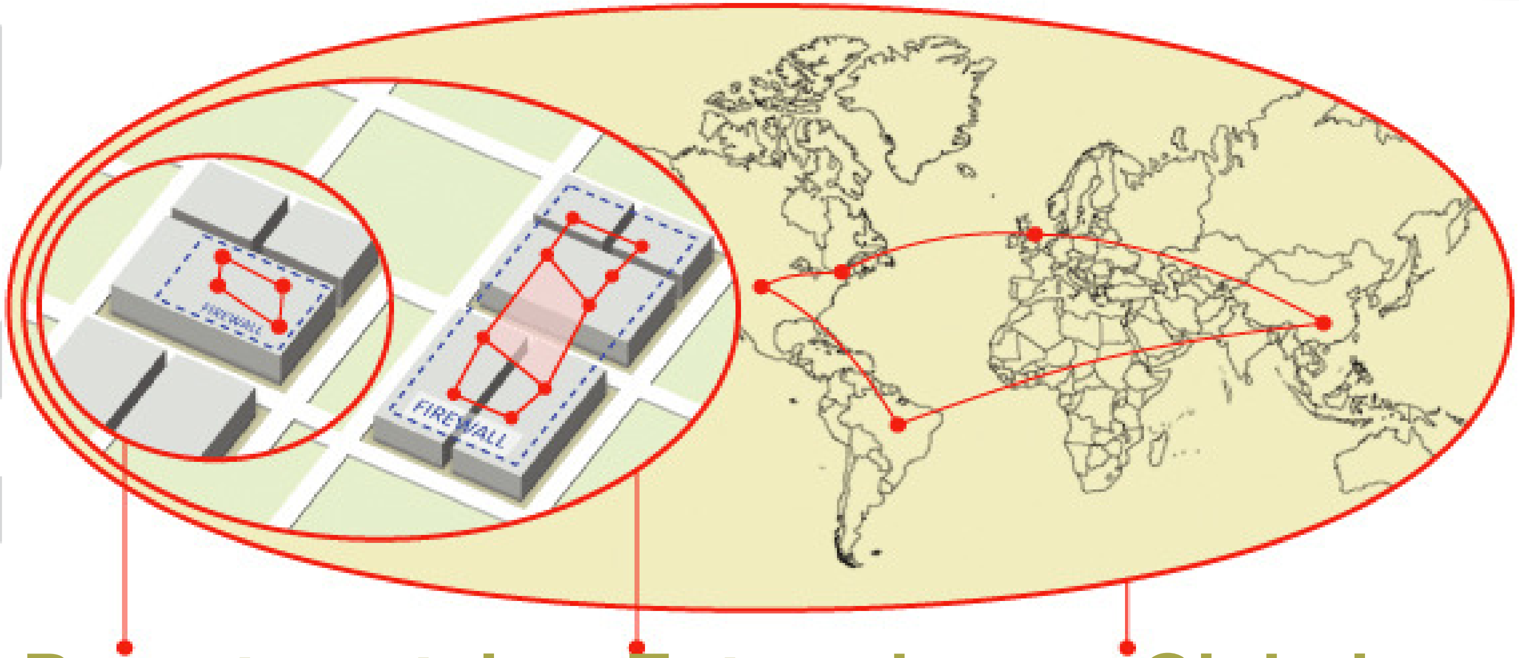
". . . coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations."

2002: Grid Checklist:

1) *coordinates resources not subject to centralized control …*
2) *… using standard, open protocols and interfaces*
3) *… to deliver nontrivial qualities of service.*

Quotes: Ian Foster, Carl Kesselman, Steve Tuecke

Grids



Departmental Grids

Enterprise Grids

Global Grids

Research is using Grids

TERAGRID

Distributed European Infrastructure for

UNICORE

Distributed European Infrastructure for Supercomputing Applications

Open Science Grid

INFN GRID

EUROGRID

Data GRID

National Grid NG SINGAPORE

Grid3 2003

PPDG

iVDgL

K*GRID

ACCESS GRID

grangenet

GRIP UK Particle Physics

DataTAG

NEESgrid

Building the National Virtual Collaboratory for Earthquake Engineering.

GriPhyN

Data Intensive Science

CERN

openlab for DataGrid applications

Developing Solutions for the Data-Intensive Science of the Large Hadron Collider

LCG

Grid5000

WestGrid

NORDUGRID

Grid Solution for Wide Area Computing and Data Handling

GridCat

ApGrid

Asia-Pacific Grid

Grid Consortium Japan

Ninf

GridLab

egee

Enabling Grids for E-science in Europe

NAREGI

超高速コンピュータ網形成プロジェクト

National Research Grid Initiative

Grid Applications

Grid Middleware

Networking

国立情報学研究所グリッド研究開発推進拠点 NII -The National Institute of Informatics

Grid.it project

Grid-Ireland

Industry is using Grids

Clouds

- IT resources provisioned outside corporate data center
- Resources accessed over the internet
- Variable cost of services
- Service oriented: SaaS, PaaS, IaaS, HaaS
- A virtual computing environment (Vmware, Xen,...)
- Build and deliver, always-on, pay-per-use IT services
- Scaling: computing, storage, database, services, and users, up and down
- Abstraction of the hardware from the service

The Cloud of Cloud Companies

Distributed
European
Infrastructure for
Supercomputing
Applications

- Amazon
- Google
- Salesforce
- Microsoft
- Sun
- IBM
- Oracle
- EMC
- Cloudera
- Cloudsoft

- Akamai
- Areti Internet
- Enki
- Fortress ITX
- Joyent
- Layered Technologies
- Rackspace
- Terremark
- Xcalibre
-
-

Distributed
European
Infrastructure for
Supercomputing
Applications

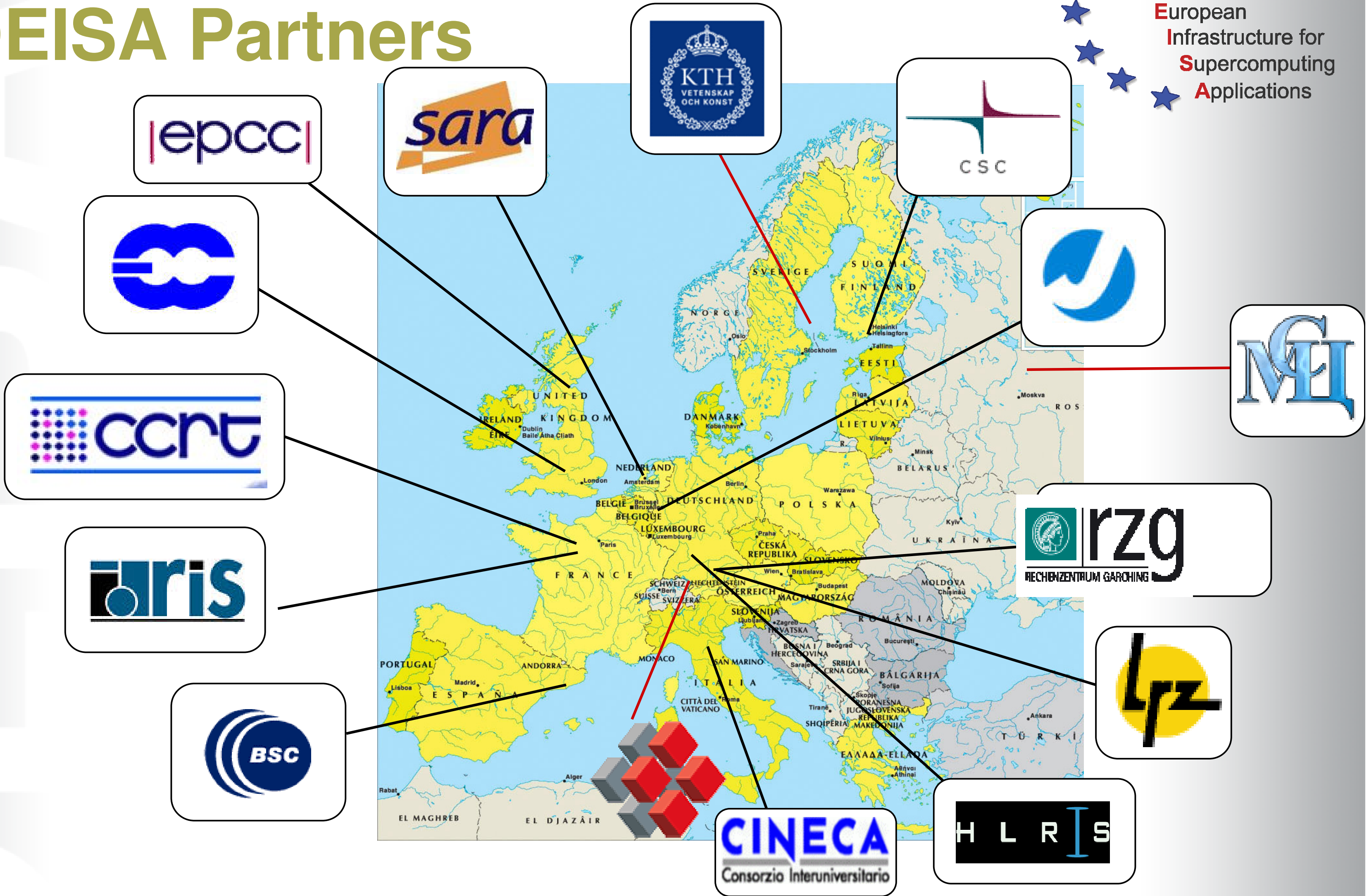
Example of an e-Infrastructure:

The DEISA Ecosystem for HPC Grand-Challenge Applications

Distributed European Infrastructure for Supercomputing Applications

DEISA Partners

Distributed European Infrastructure for Supercomputing Applications



DEISA1: May 1st, 2004 – April 30th, 2008

DEISA2: May 1st, 2008 – April 30th, 2011

DEISA: Vision - Mission - Strategy

Distributed
European
Infrastructure for
Supercomputing
Applications

Vision:

Establishing persistent European **HPC ecosystem** integrating national Tier-1 (Tflop/s) centres and the new European Tier-0 (Pflop/s) centres

Mission:

Enhance Europe's capability in computing and science by **integrating** most powerful **supercomputers** into a European HPC e-infrastructure

Build European Supercomputing **Service** on top of **existing** national services, based on the deployment and operation of a persistent, **production** quality, distributed supercomputing environment with continental scope

SEVENTH FRAMEWORK PROGRAMME
RI-222919

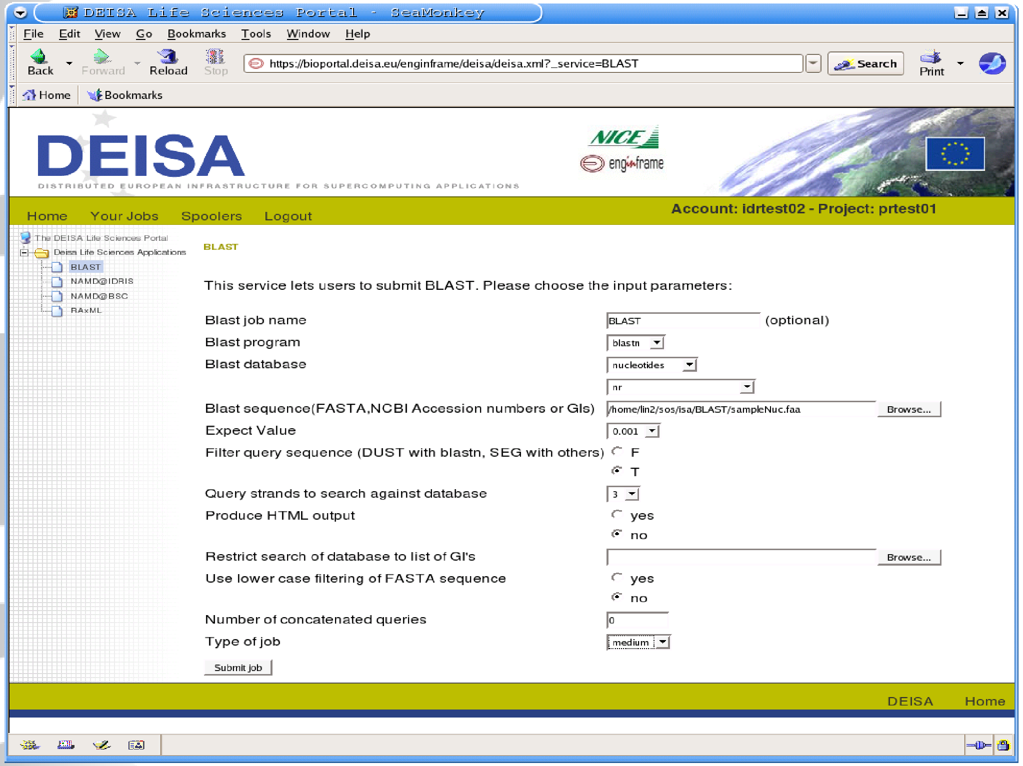
DEISA Service Layers

Multiple ways to access	Workflow managemnt	Common production environmnt	*Presen-tation layer*
Single monitor system	Job rerouting	Co-reservation and co-allocation	*Job manag. layer and monitor.*
Data staging tools	Data transfer tools	WAN shared File system	*Data manag. layer*
Unified AAA	DEISA Sites	Network connectivity	*Network and AAA layers*

One Example of Virtual Communities: Joint Research Activity "Life Sciences"

Distributed
European
Infrastructure for
Supercomputing
Applications

The DEISA Life Science Portal



Joint Research Activity (JRA)

Promoting parallel apps in the life science community

Running big simulations on DEISA infrastructure that couldn't be done locally

Providing ease of access to resources

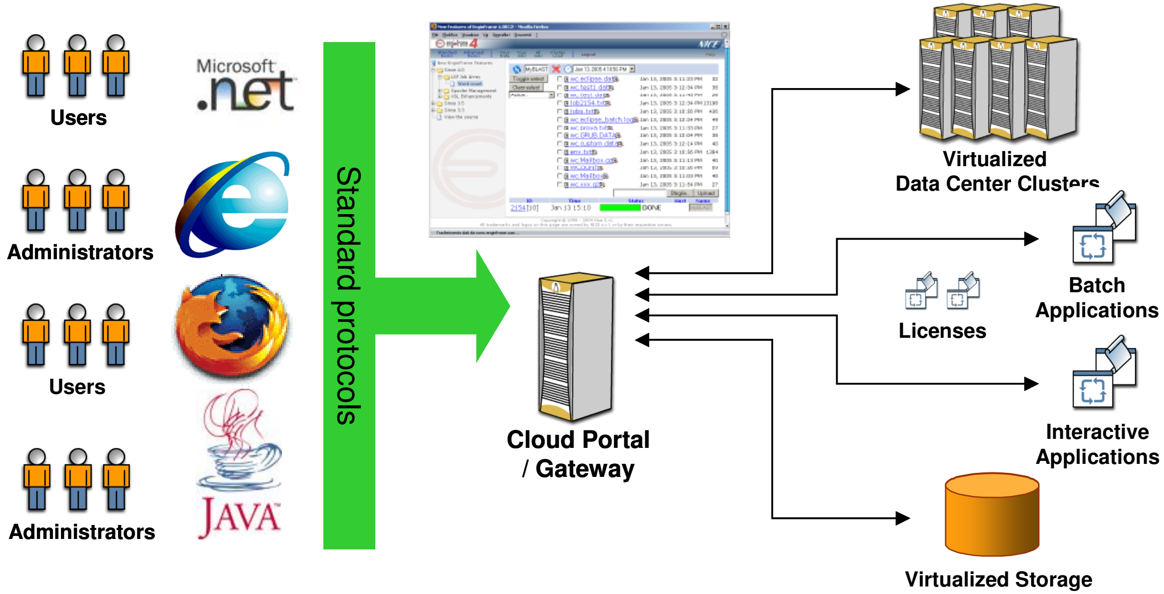
Application support for life science portal

Cluster/Grid/Cloud Portal

Example: NICE EnginFrame

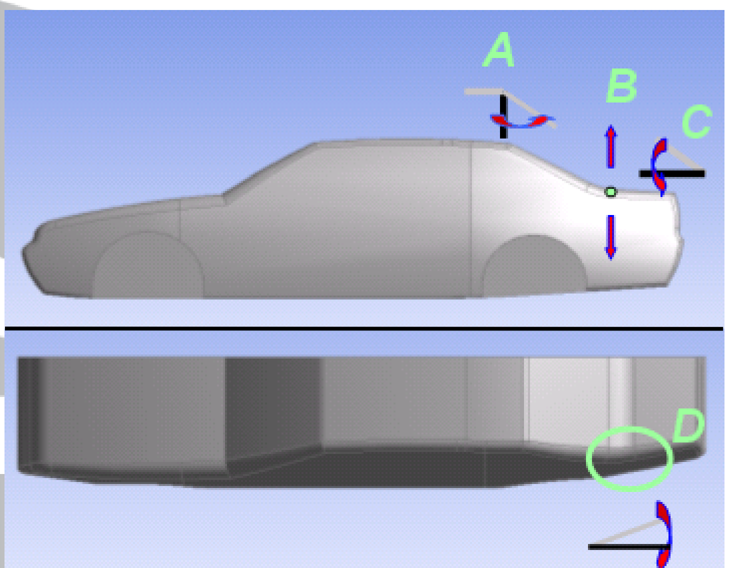
Distributed European Infrastructure for Supercomputing Applications

Provides remote, interactive, transparent, and secure access to applications and data on your corporate Intranet or Internet, or in the Cloud.

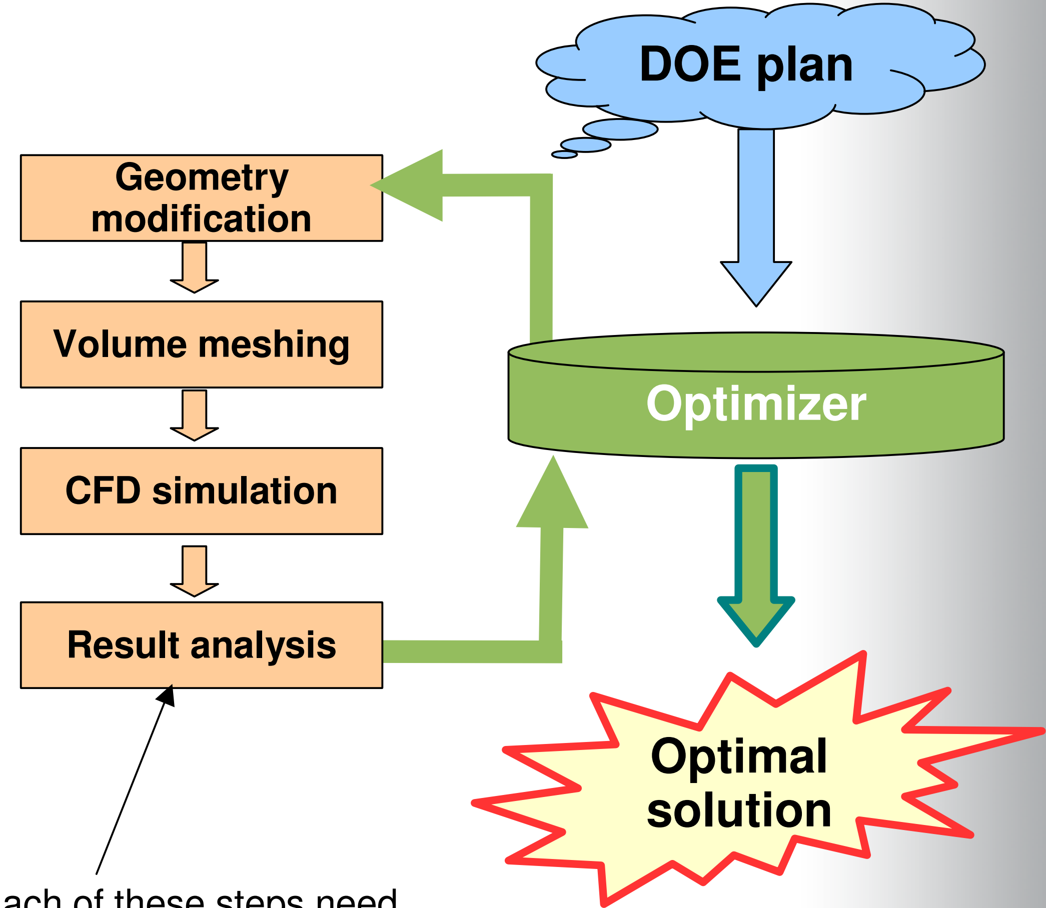


Users and administrators can access and control computing resources via an intuitive and standard Web interface virtually anywhere using a standard Web browser.

Aerodynamic Shape Optimization



- 4 parameters to be optimized
- cubic face centered DOE
- 25 cases+16 extra cases for error estim.
- polynomial response function
- 70 hours wall clock time on 64 cpus



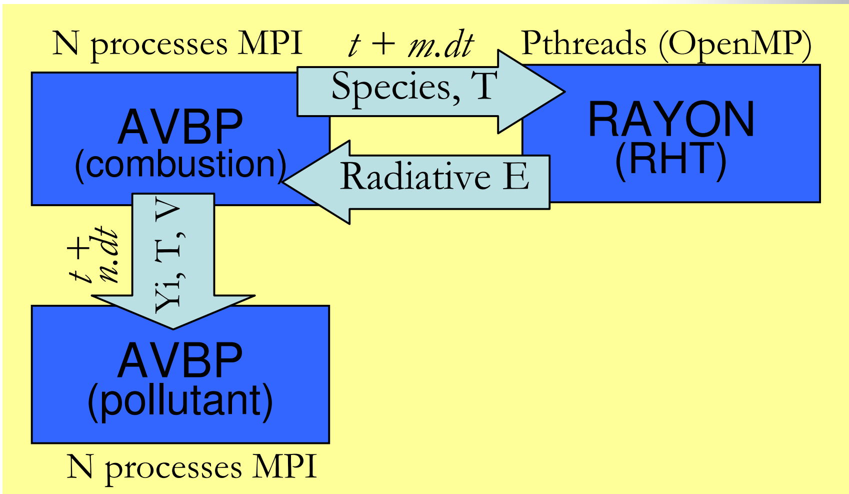
Each of these steps need to be fully automated and controlled by the optimizer

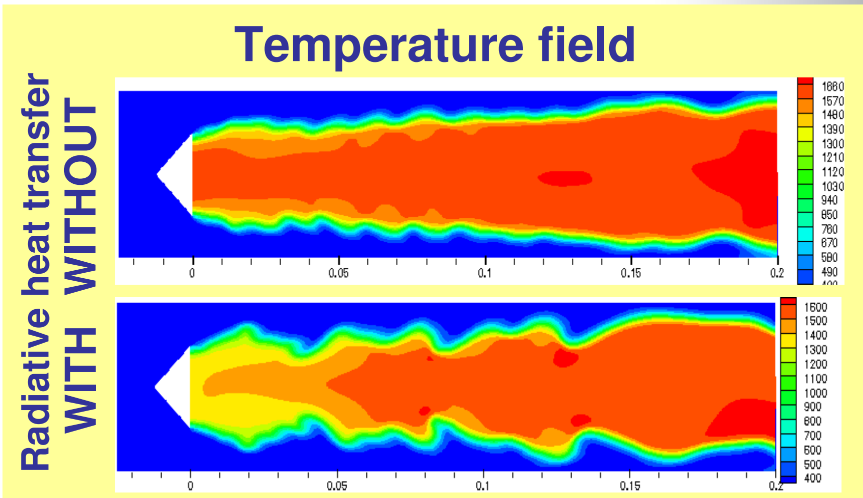
RI-222919

Combustion / Radiation

Distributed European Infrastructure for Supercomputing Applications

- Study the impact of radiative heat transfer (RHT) on the combustion process (2D)
- Couple combustion (AVBP), the RHT (Rayon) codes and the pollutant formation (AVBP)
- Parallelization of the Rayon code and improvement of the coupling part
- Load balancing issue
- 3D extension proposed to DECI and accepted



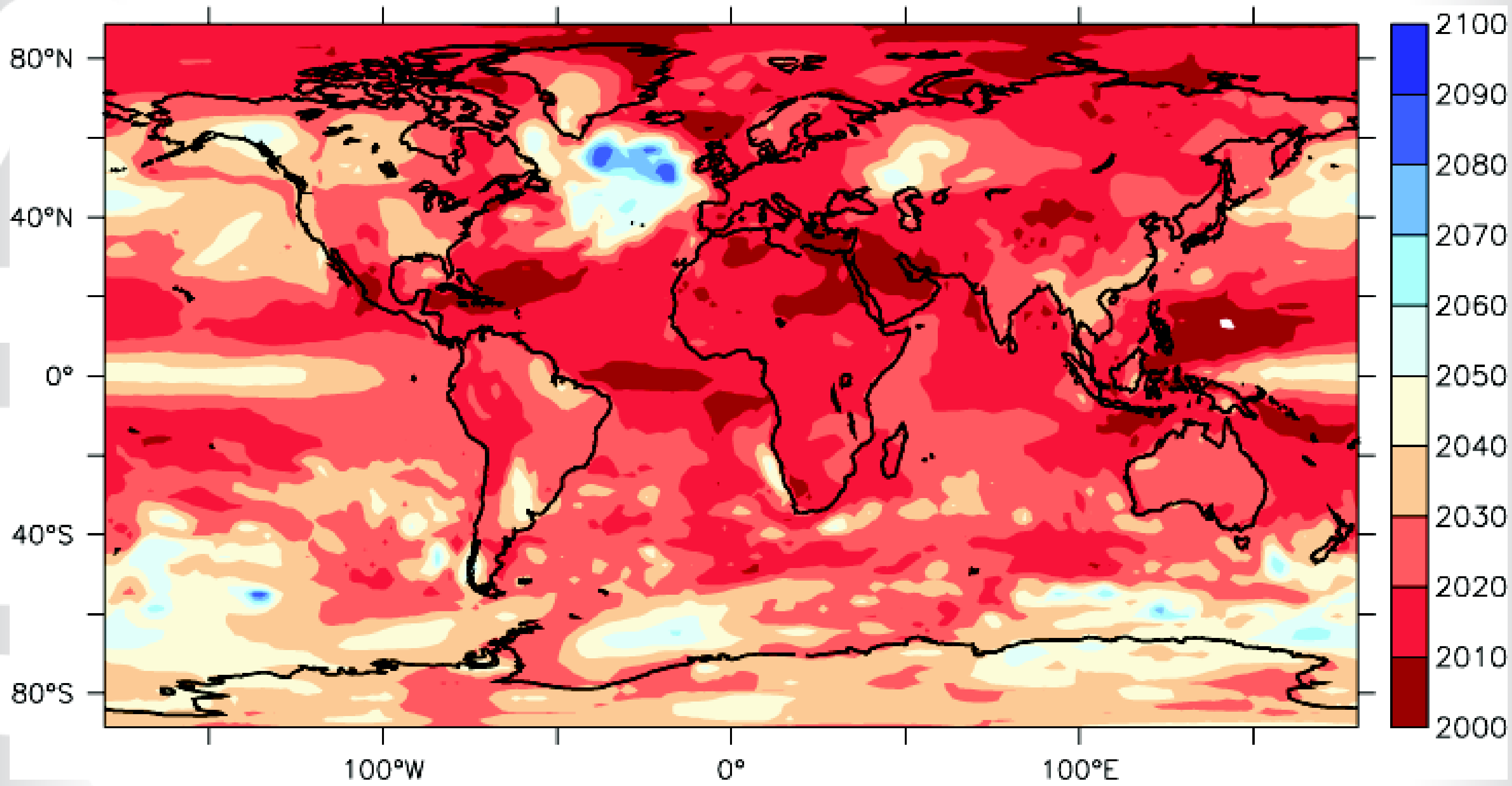


RI-222919

Climate Research
Statistics of Climate Variability

Distributed
European
Infrastructure for
Supercomputing
Applications

Project to study climate trends, each 50 TB output data

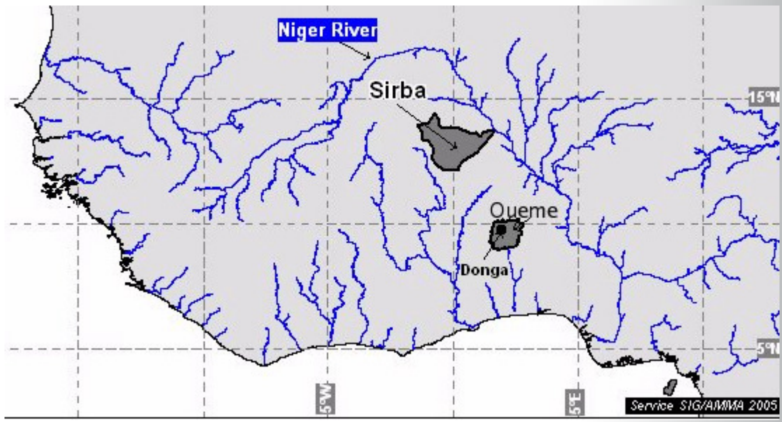


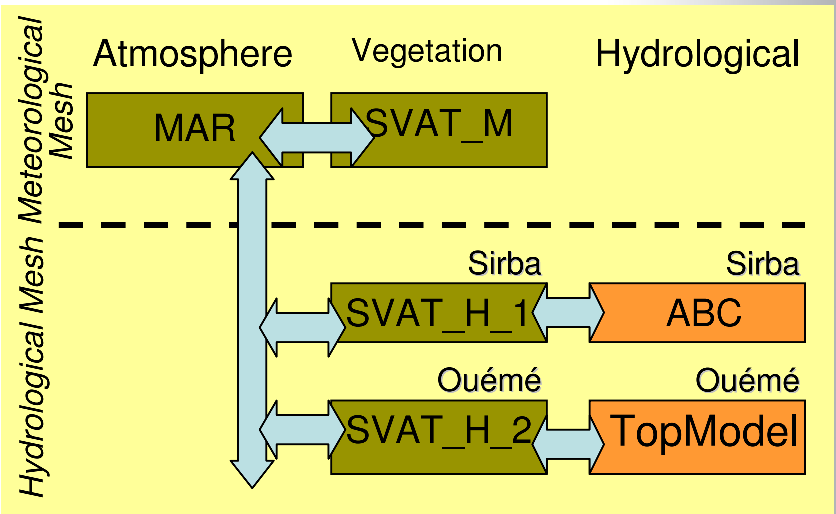
H. Dijkstra, U Utrecht, and W. Hazeleger, KNMI

RI-222919

Environmental Application

- Study the impact of water cycles of the hydrological and vegetation models on climate models
- Coupling area in West Africa
- Best performances with a vector and scalar platform
- Improve extensibility of the architecture and the coupling part
- AMMA project, PhD thesis, 2 publ. and 2 comms.



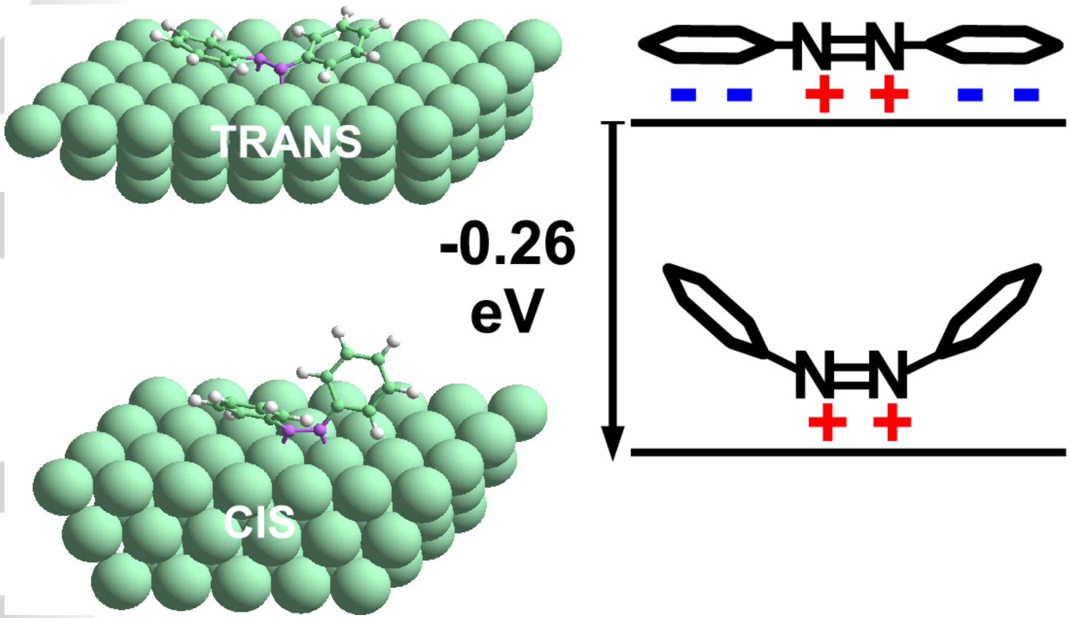


RI-222919

Materials Science

First-principles statistical mechanics for molecular switches at surfaces (MolSwitch)

Azobenzene on copper, silver and gold surfaces



Controlled reversible switching should be possible on Ag surfaces

Courtesy: K. Reuter, FHI

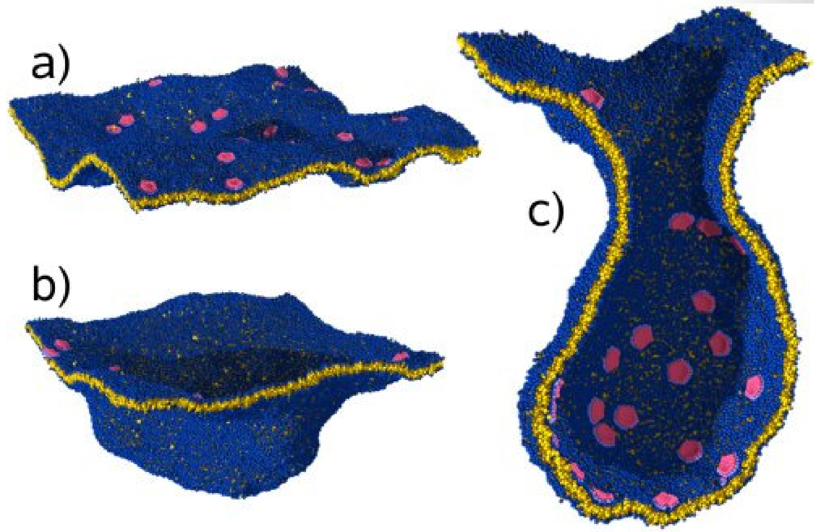
Polymer Research

Cover story of Nature - May 24, 2007

Curvy membranes make proteins attractive

For almost two decades, physicists have been on the track of membrane mediated interactions. Simulations in DEISA have now revealed that curvy membranes make proteins attractive

Nature 447 (2007), 461-464



a) proteins (red) adhere on a membrane (blue/yellow) and locally bend it;

b) this triggers a growing invagination.

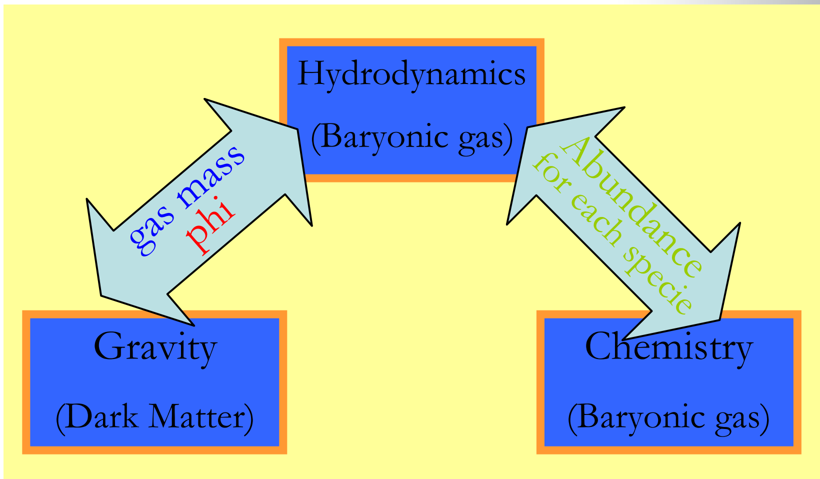
c) cross-section through an almost complete vesicle

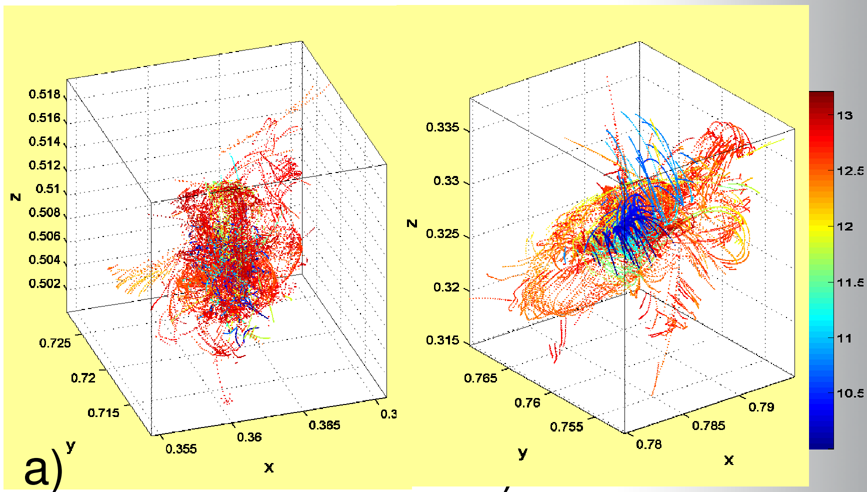
B. J. Reynwar et al.: ***Aggregation and vesiculation of membrane proteins by curvature mediated interactions***, NATURE Vol 447|24 May 2007| doi:10.1038/nature05840

Cosmology Project

Distributed European Infrastructure for Supercomputing Applications

- Study galaxy formation in cosmology
- Physics / modules: Gravitation, Hydrodynamics, Chemistry
- Best performance on heterogeneous platforms
- Load balancing issue and improvement of the coupling part
- Proposed to DECI





RI-222919

Next-Generation e-Infrastructures for

ACHI: Advances in Computer-Human Interactions
eKNOW: Information, Process, and Knowledge Management
eL & mL: Mobile, Hybrid, and On-line Learning
eTELEMED: Health, Telemedicine, and Social Medicine
GEOWS: Advanced Geographic Information Systems & Web Services
ICDS: Digital Society
ICQNM: Quantum, Nano, and Micro Technologies

Connected: anyone, anywhere, anytime, any device

- Integration of new devices, data and information sources
- Cell phones, PDAs, smart sensors, sensor arrays, health monitors
- Devices embedded in cars, engines, roads, bridges, clothes,...
- Huge amount of data for real-time analysis
- Policies, grid economy, to maintain stability and efficiency
- Support organizational and societal structures, to bridge political and social boundaries . . .

Example:
e-Learning on e-Infrastructures

Distributed
European
Infrastructure for
Supercomputing
Applications

The Challenges:

World-wide data and knowledge explosion

We need more scientists and engineers, but not enough students are interested in science

Schools and teachers are not prepared

The Solution:

New ways of teaching and learning for our digital natives

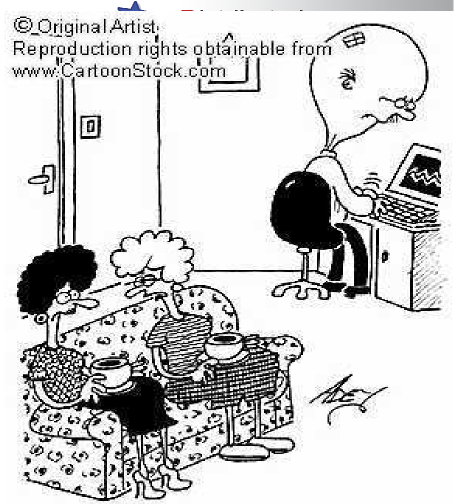
e-Infrastructures for enriched learning

Working with didactic and pedagogic experts

The Prototype:

e-School, interactive science laboratory, the digital sand-box for life-long learning in the sciences

Today’s Education Challenges



I’ve tried to tell Bill he’s overloading himself with too much information

- **Information** & knowledge growing exponentially
- **Teaching** methods + materials do not keep pace
- **Learning** is too passive and static, life is highly active and dynamic

  Prof. Srivathsan, India: „...education today is not learning centric, it is exams centric.“
- **Students** become de-motivated and lack creativity
- **e-Learning** environments just scratching the surface

We need <u>100Ks</u> of new jobs in science and engineering

We have to focus on K-12 students and their teachers

e-School Prototype

A **Virtual Laboratory** based on an e-Infrastructure and a distributed digital repository for science and engineering applications for students and educators

Bridging the Chasm between

Education  **Science**

Vision: e-School Science Collaboratory for a better learning experience

- **Inter-active** learning tools for creative students (edutainment)…
- …**same tools** engineers & scientists are using in the 21st century
- Edu portal provides seamless access to **virtual laboratory**
- 100s of real-world computer simulations available for all ages
- On dynamic, shared, remote resources, **at your finger tip**

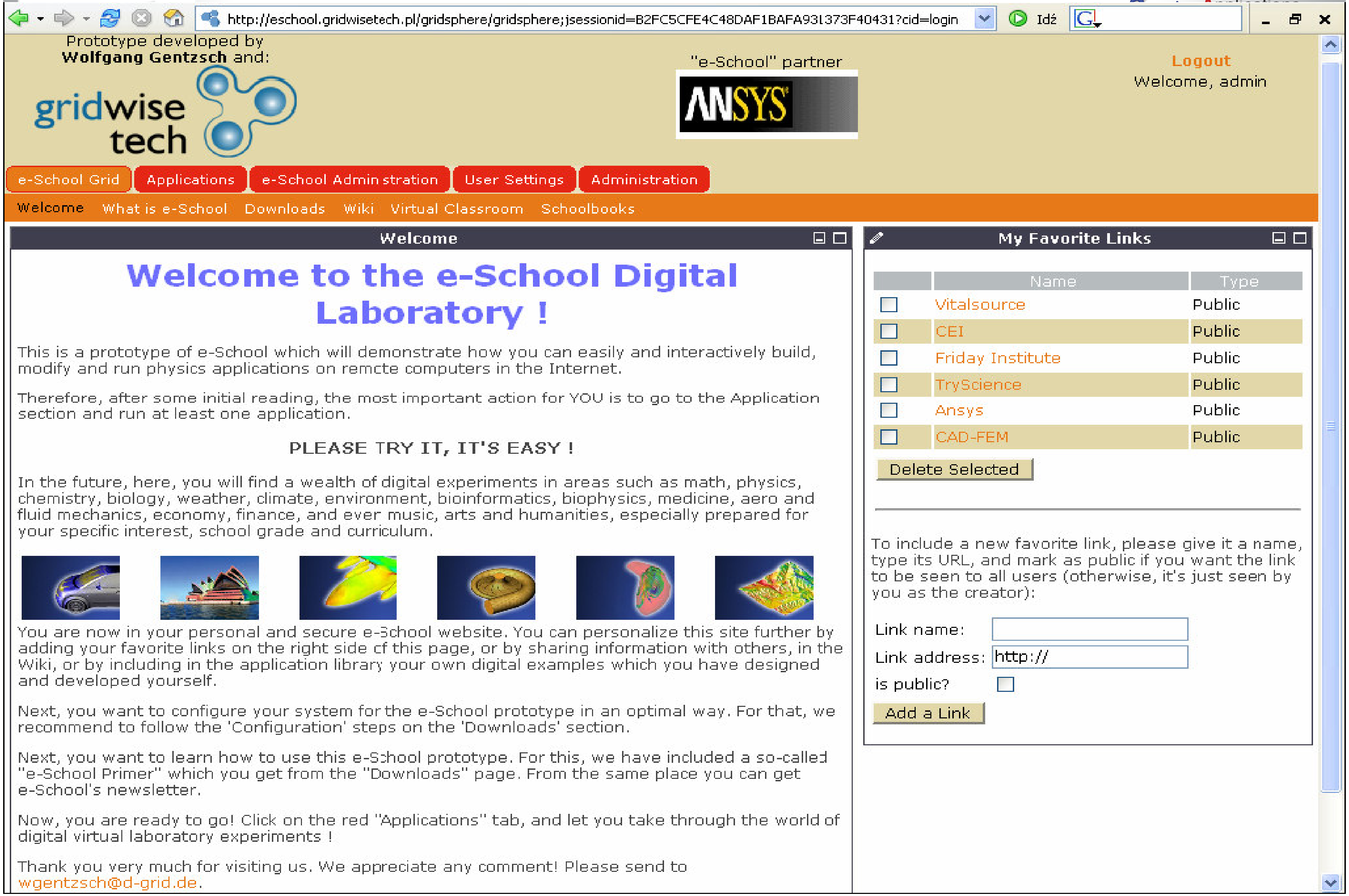
e-School: empowering education

- Learning by doing -

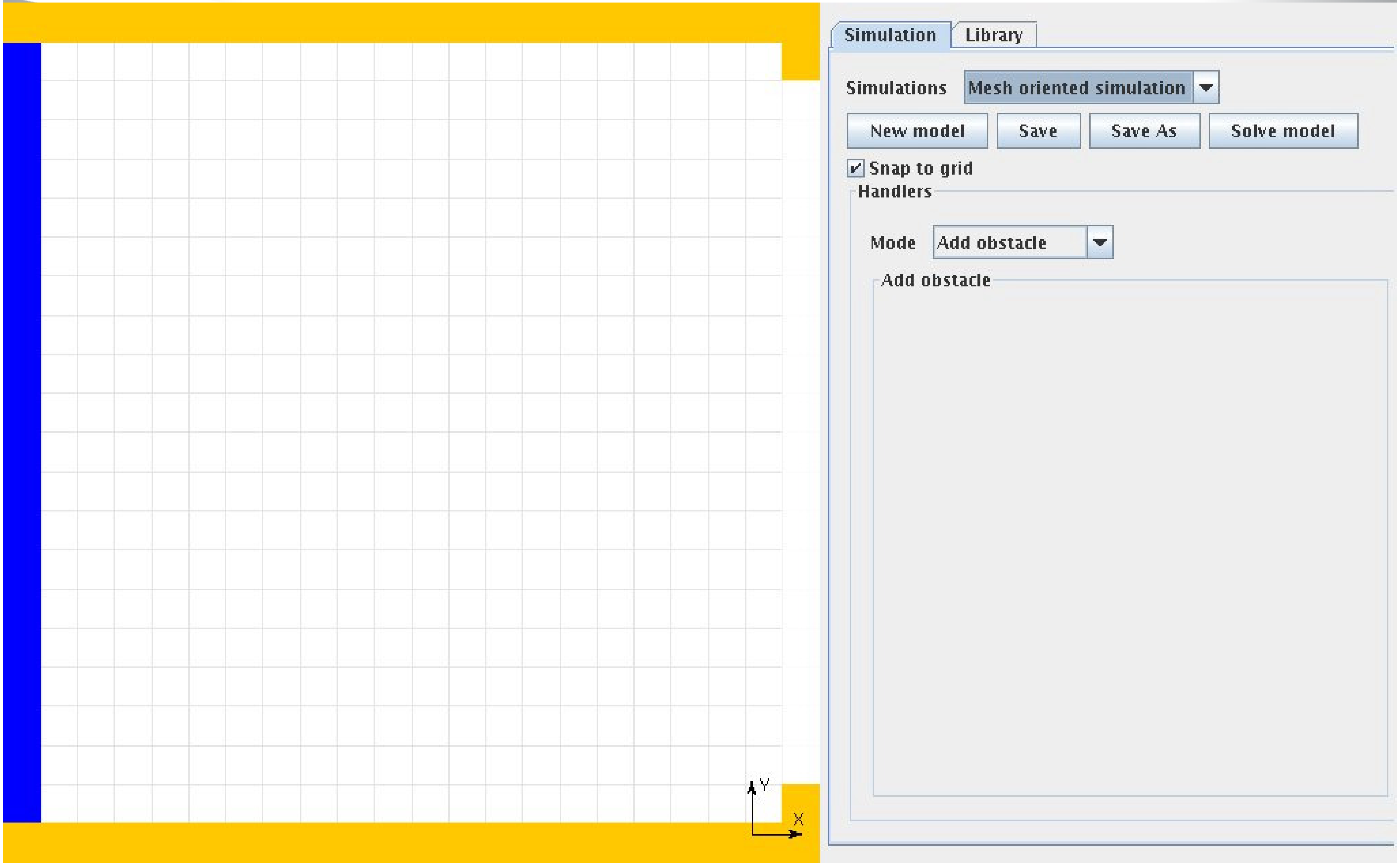
e-School Prototype Website

http://eschool.gridwisetech.pl

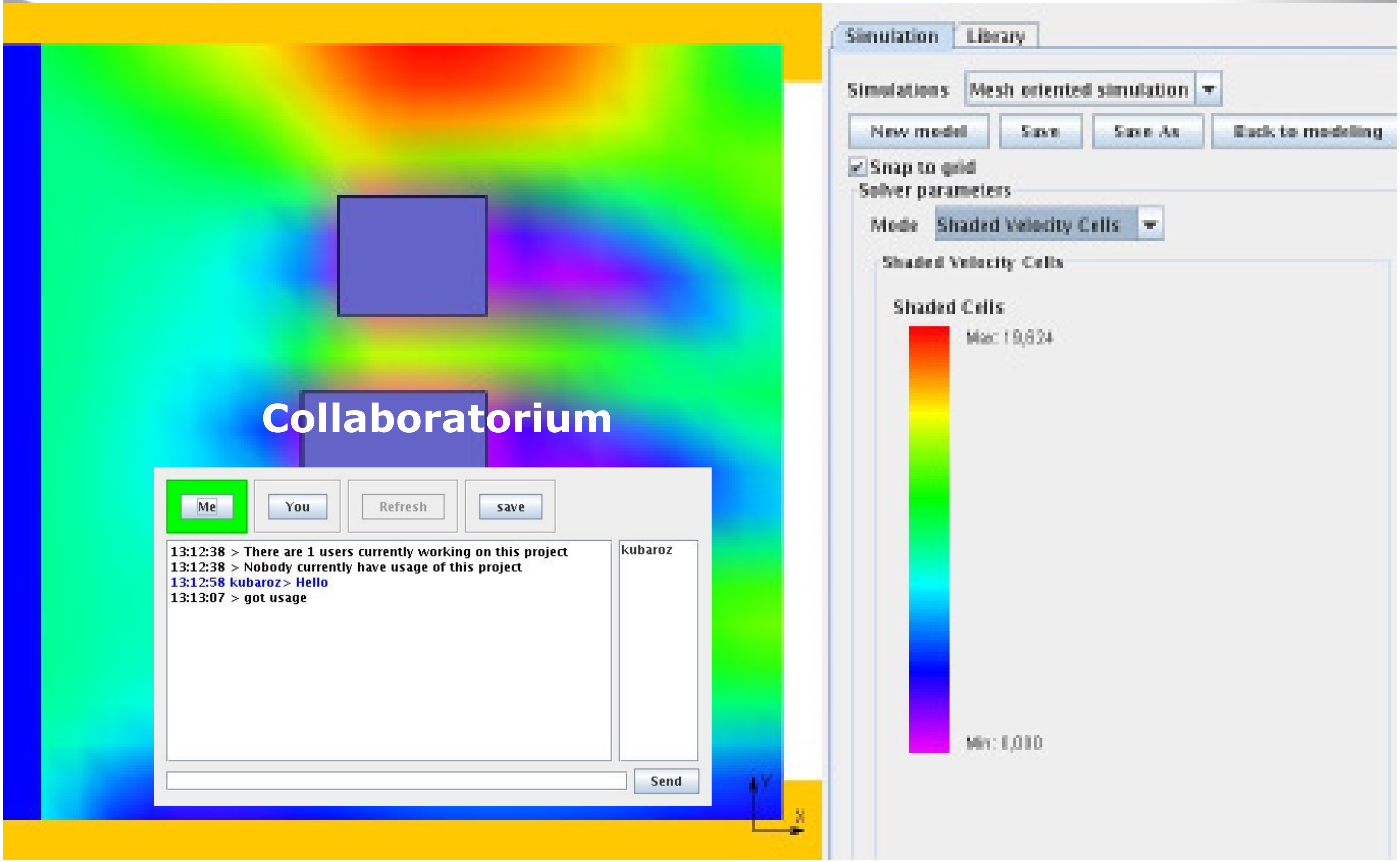
Distributed
European
Infrastructure for
Supercomputing



e-School: Your personal workspace



Example: interactive real-time fluid flow



Distance Learning on e-Infrastructures

Distance Learning...

- Independent of time and space
- Self-paced learning
- Teacher-independent learning
- Deductive science education
- Mostly single-learner envirnmt.
- Linear inter-reactivity, at best
- Mostly static and repetitive

=> Improved (but similar to) class-room learning

...on e-Infrastructures

- Independent of time and space
- Self-paced learning
- Teacher-independent learning
- Inquiry-based science education
- Allows for collaborating groups
- Fully nonlinear interactivity
- Highly dynamic and nonlinear
- Allows for complex simulations, data processing, and visualization
- Students and teachers become more creative, motivated and committed

=> Paradigm Shift

Distributed
European
Infrastructure for
Supercomputing
Applications

Read more in:

Thank You!

GRACIAS POR SU ATENCIÓN

Gentzsch@rzg.mpg.de