Grids and EGEE are not just for High Energy Physicists

Richard Hopkins, National e-Science Centre

June 29, 2005

Overview

- **Goals - An appreciation of**
  - the range of potential (non-physics) Grid application areas
  - the process by which new application areas are integrated into EGEE as new VOs
- **Outline**
  - Biomed – the other pilot application
  - Some other potential application areas –
    - Earth observation
    - Weather Forecasting
    - Engineering
    - e-Research and beyond
  - The process for new VO's
  - The up-coming VOs –
    - Computational Chemistry
    - Earth Science
    - Astrophysics

Acknowlegements – mainly a talk prepared by Favid Fergusson, NeSC

The characteristics of biomedical pilot applications (vs HEP)

Enabling Grids for E-sciencE

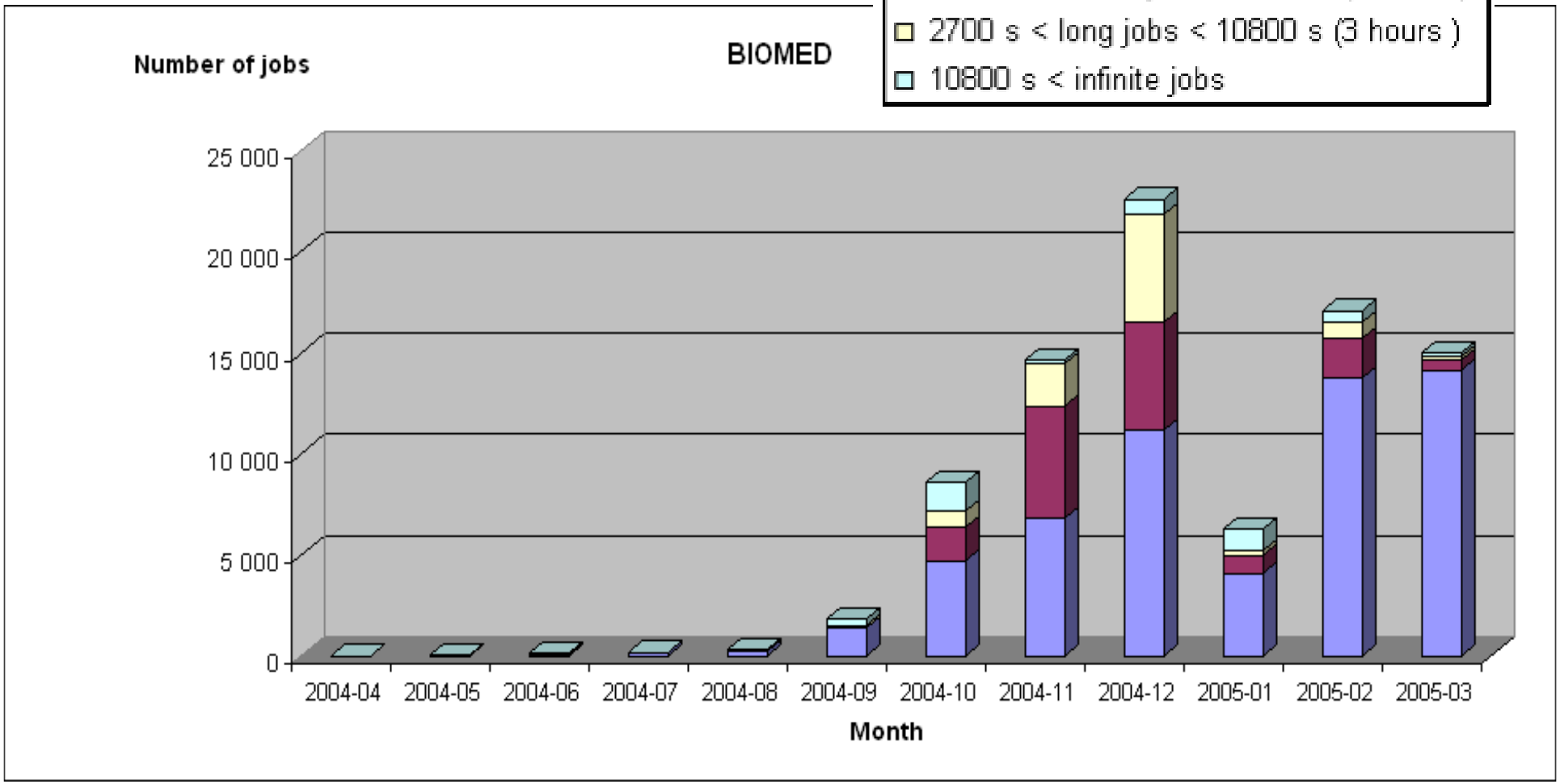
- **Prototype level at project day 1**
  - HEP very large scale from day 1
- **VO was created after the project kicked-off**
  - HEP -Virtual Organisations were already set up at project day 1
- **Very decentralized: application developers use the grid at their own pace**
  - HEP - Very centralized: jobs are sent in a very organized way
- **Very demanding on services**
    - Compute intensive applications
    - Applications requiring large amounts of short jobs
    - Need for interactivity or guaranteed response time
  - HEP – Primarily requires “Data Distribution grid” the data challenges
- **Resources were focused on the deployment of large scale applications on LCG-2**
  - HEP – data challenges deployed on several grids

- **Decentralized usage of the infrastructure highlights different weaknesses from the more centralized HEP data challenges**
  - Integration of Biomed VO used to identify issues relevant to all VOs to be deployed during EGEE lifetime
  - Generally an application is some combination of HEP/Biomed features

Status of Biomedical VO

Infrastructure usage

- **JRA2 statistics**
  - ~15Kjobs per month



Biomedical applications

- 3 batch-oriented applications ported on LCG2
  - SiMRI3D: medical image simulation
  - xmipp_MLRefine: molecular structure analysis
  - GATE: radiotherapy planning
- 3 high throughput applications ported on LCG2
  - CDSS: clinical decision support system
  - GPS@: bioinformatics portal (multiple short jobs)
  - gPTM3D: radiology images analysis (interactivity)
- Recent Additions
  - xmipp_ML_refine: Macromolecular 3D structure analysis (CNB)
  - xmipp_multiple_CTFs : Electronmicroscopic images CTF calculation (CNB)
  - GridGRAMM: Molecular Docking web (CNB)
  - GROCK: Mass screenings of molecular interaction (CNB
  - Mammogrid: Mammograms analysis (EU project)
  - SPLATCHE: Genome evolution modeling (U. Berne/WHO)

Evolution of biomedical applications

- **Growing interest of the biomedical community**
  - Partners involved proposing new applications
  - New application proposals (in various health-related areas)
  - Enlargement of the biomedical community (drug discovery)

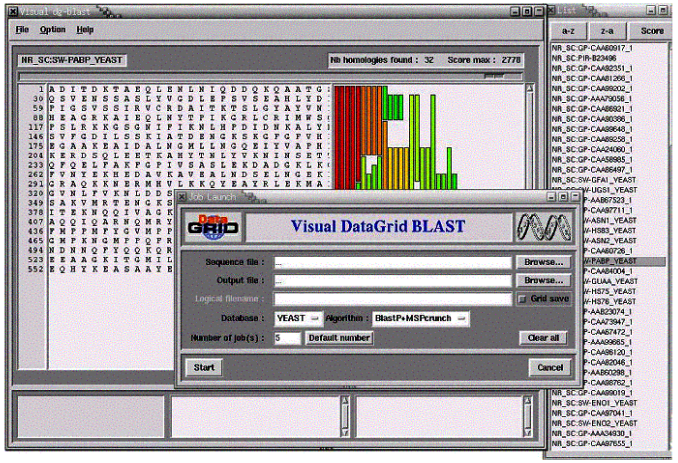
- **Growing scale of the applications**
  - Progressive migration from prototypes to pre-production services for some applications
  - Increase in scale (volume of data and number of CPU hours)

- **Towards pre-production**
  - Several initiatives to build user-friendly portals and interfaces to existing applications in order to open to an end-users community

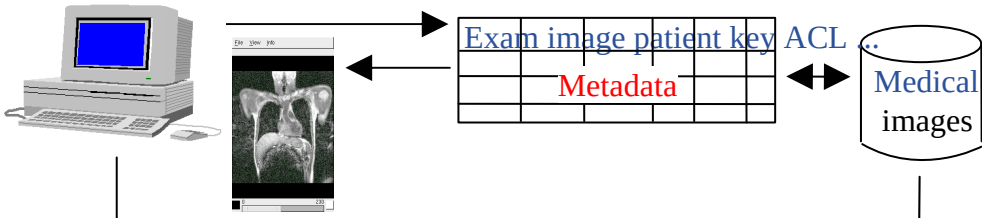
Bio-medicine applications

- Bio-informatics
  - **Phylogenetics ***
  - **Search for primers ***
  - **Statistical genetics**
  - **Bio-informatics web portal**
  - **Parasitology ***
  - **Data-mining on DNA chips**
  - **Geometrical protein comparison**
- Medical imaging
  - **MR image simulation**
  - **Medical data and metadata management ***
  - **Mammographies analysis ****
  - **Simulation platform for PET/SPECT ****

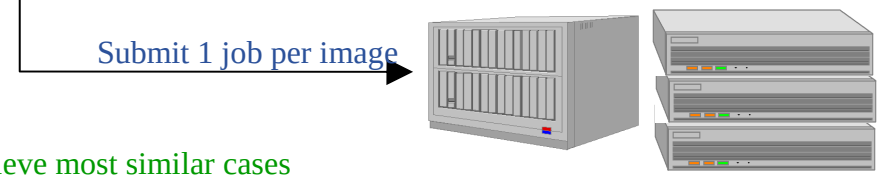
Applications deployed *
Applications tested **
Applications under preparation



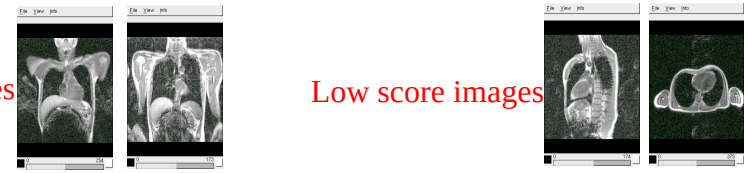
1. Query the medical image database and retrieve a patient image



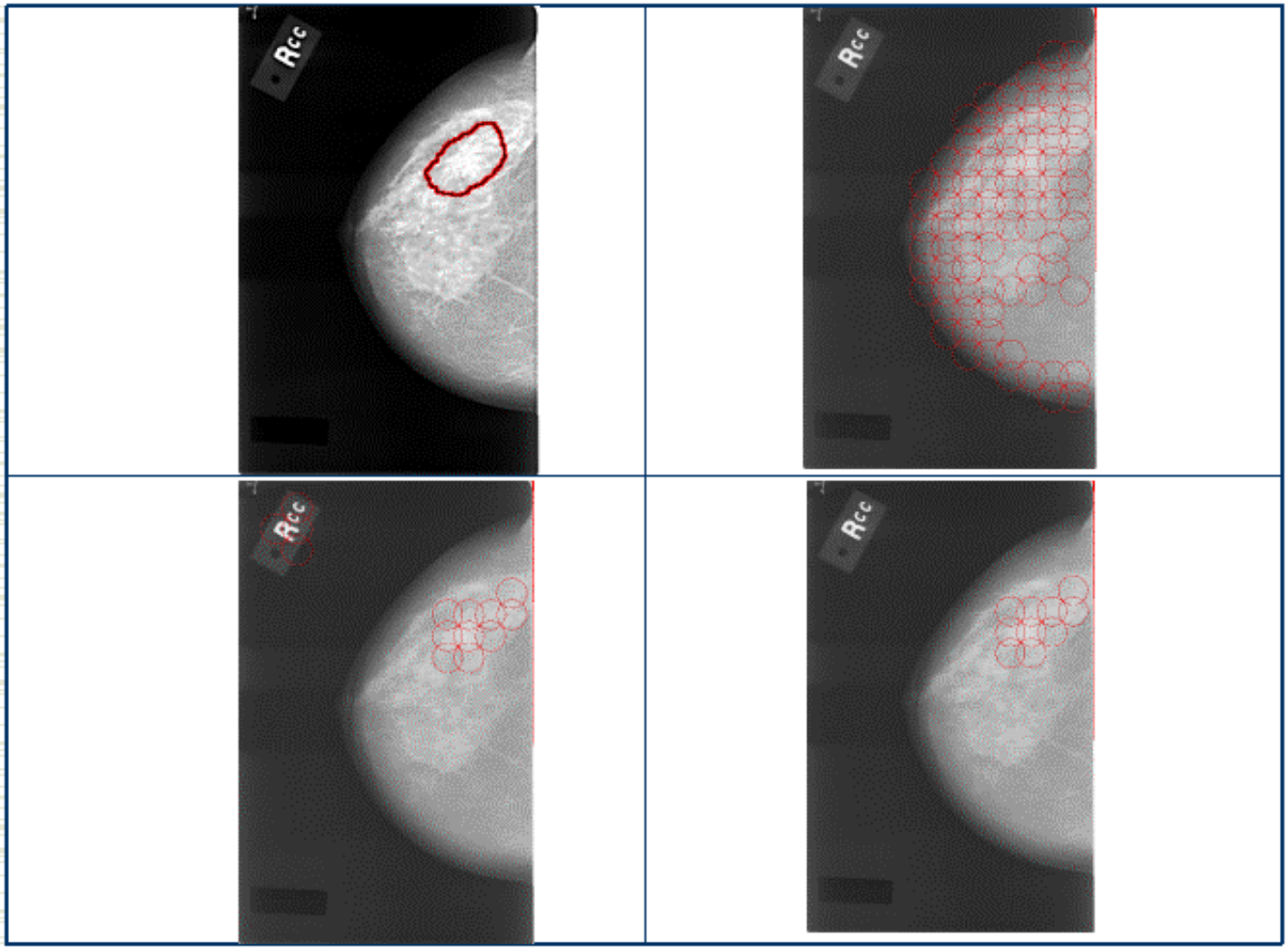
2. Compute similarity measures over the database images



3. Retrieve most similar cases



Bio-medicine applications

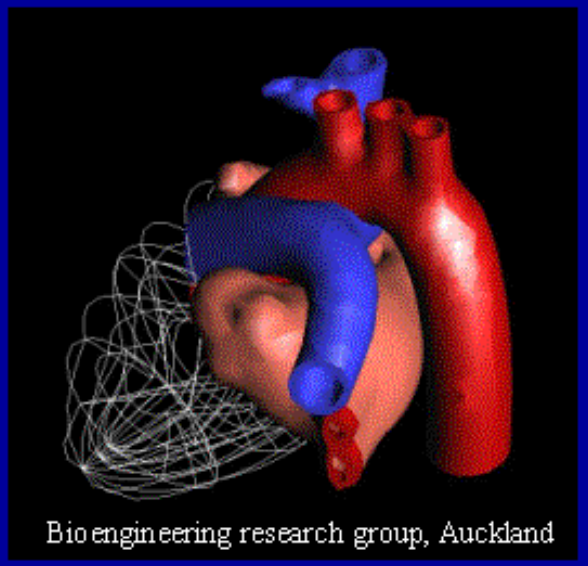


September 21st, 2001 *Datagrid Meeting, Lyon* 15

Bio-medicine applications

3.3 Heart Modeling

- Objectives: modeling heart anatomy, dynamics and physiology for heart image processing
  - bio-mecanical model
  - electrical model
  - very complex structure
  - biological scale out of range



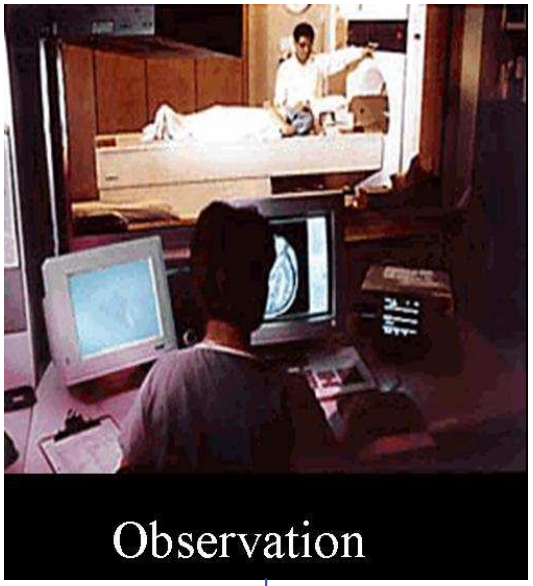
Bioengineering research group, Auckland

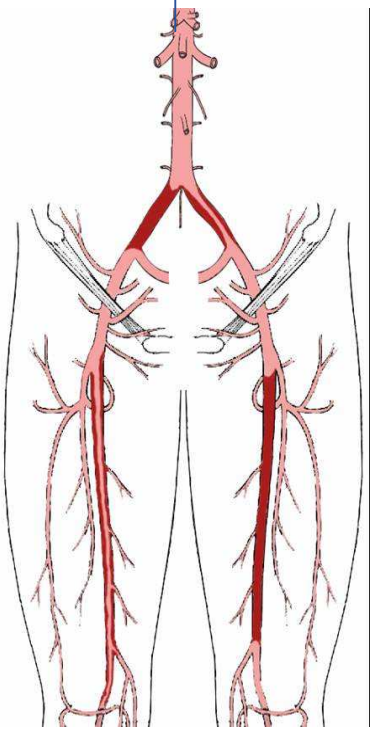
- Finite Element modeling
  - elements oriented in heart fibers direction: fine resolution
  - electrical propagation model based on bidomain theory
  - 4D model (3D+T)

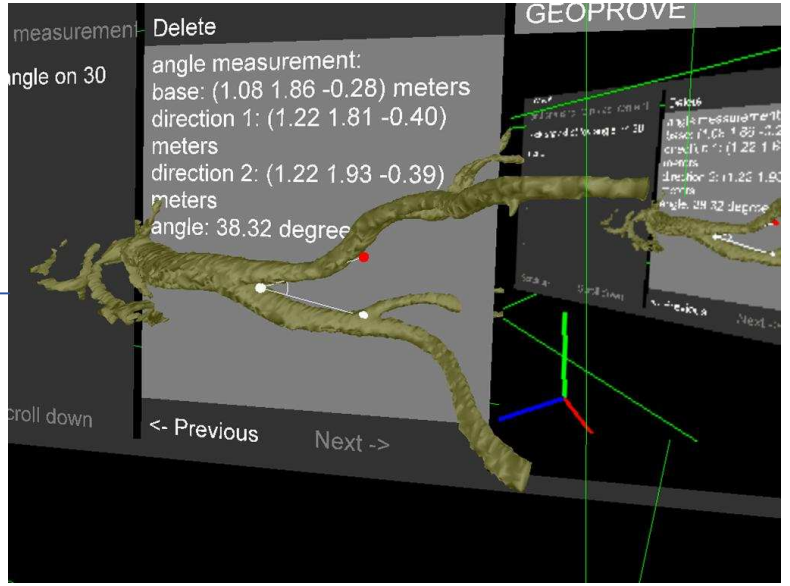
Creatis

GGF1 - DataGrid WP 10 - March 2001

Bio-medicine applications

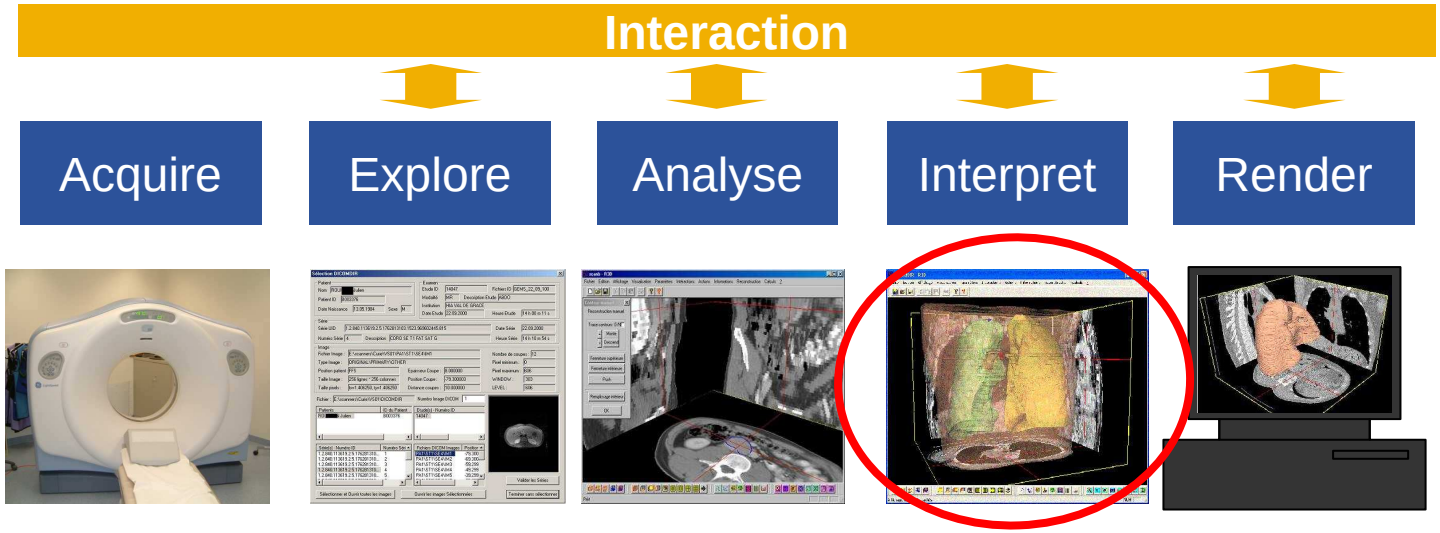






gPTM3D : Grid-Enabling Interactive Medical Analysis

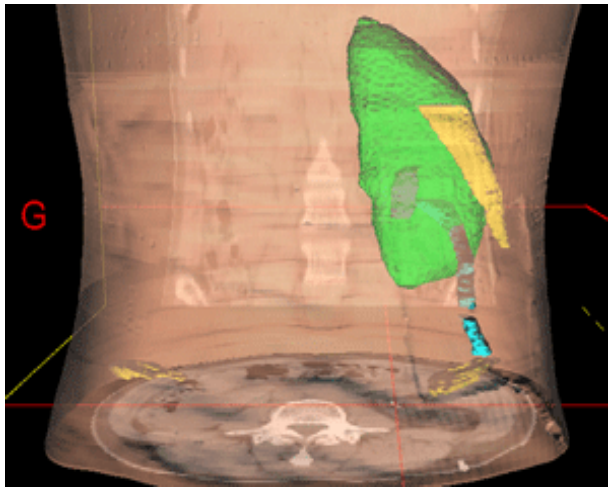
PET – Positron Emission Tomography



Construction of model has High Computational requirements

Use case

Planning percutaneous nephrolithotomy – under-skin kidney stones



Feedback to LCG-2 middleware developers and infrastructure

- **Feed-back from Biomed applications**
  - Very significant exchanges related to the set-up of the biomed VO and the deployment of relevant service
- **Very decentralized: application developers use the grid at their own pace**
- **Very demanding on services**
  - Compute intensive applications
  - Applications requiring large amounts of short jobs
  - Need for interactivity or guaranteed response time
    - *Request to use MPI*
- **Whereas HEP is primarilly Data Distribution**
- **Generally an application is some combination of HEP/Biomed features**

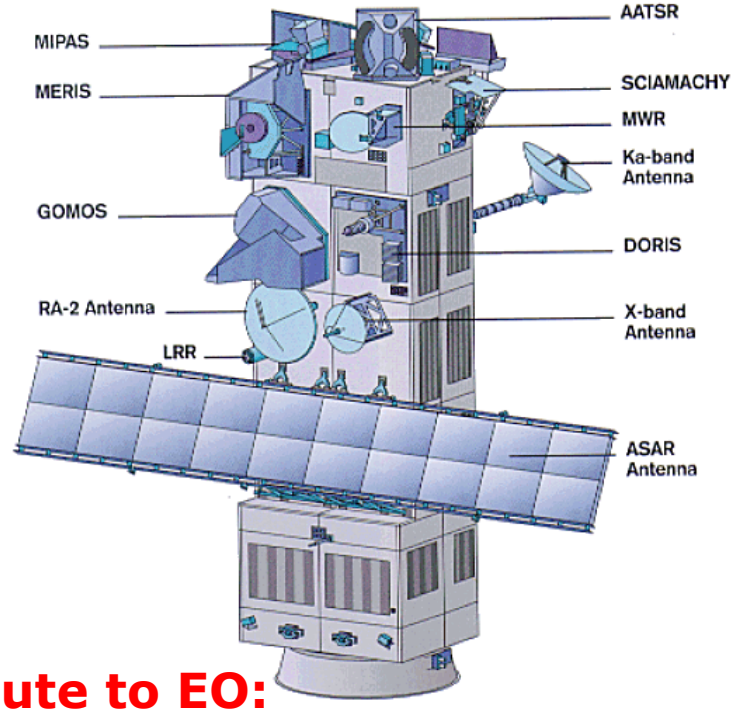
SOME OTHER POTENTIAL APPLICATION AREAS

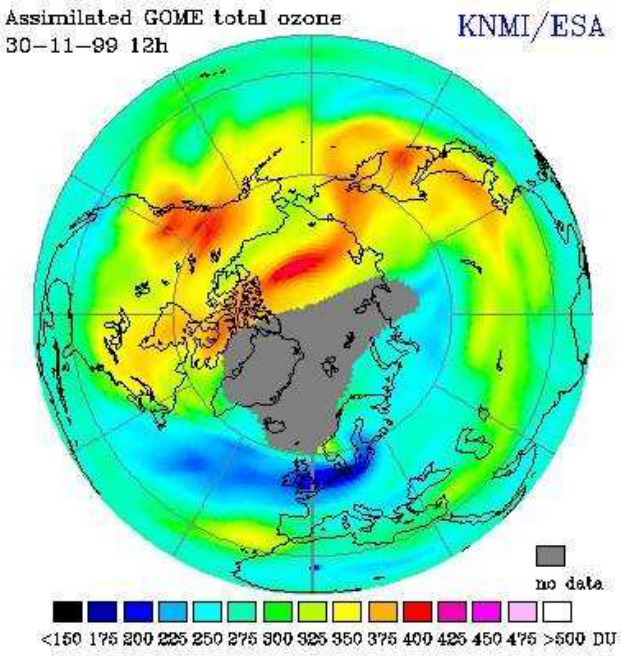
- **Goals - An appreciation of**
  - the range of potential (non-physics) Grid application areas
  - the process by which new application areas are integrated into EGEE as new VOs
- **Outline**
  - Biomed – the other pilot application
  - **Some other potential application areas –**
    - **Earth observation**
    - **Weather Forecasting**
    - **Engineering**
    - **e-Research and beyond**
  - The process for new VO's
  - The up-coming VOs –
    - Computational Chemistry
    - Earth Science
    - Astrophysics

Earth observation applications

ESA missions:

- about 100 Gbytes of data per day (ERS 1/2)
- 500 Gbytes, for the ENVISAT mission (2002).





Grid contribute to EO:

- enhance the ability to access high level products
- allow reprocessing of large historical archives
- improve Earth science complex applications (data fusion, data mining, modelling …)

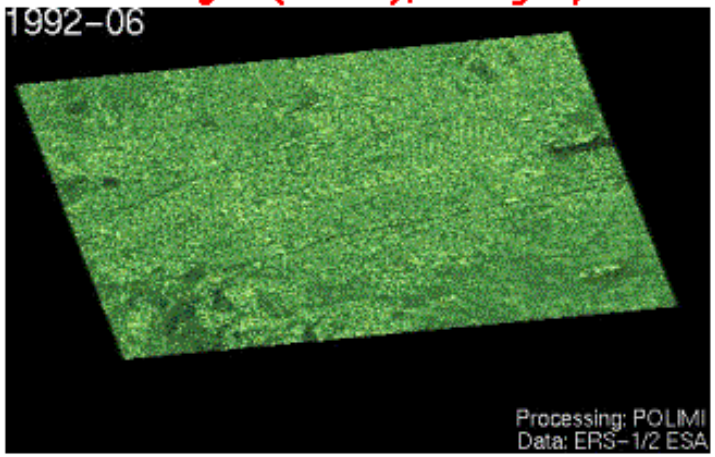
Roberto Barbera

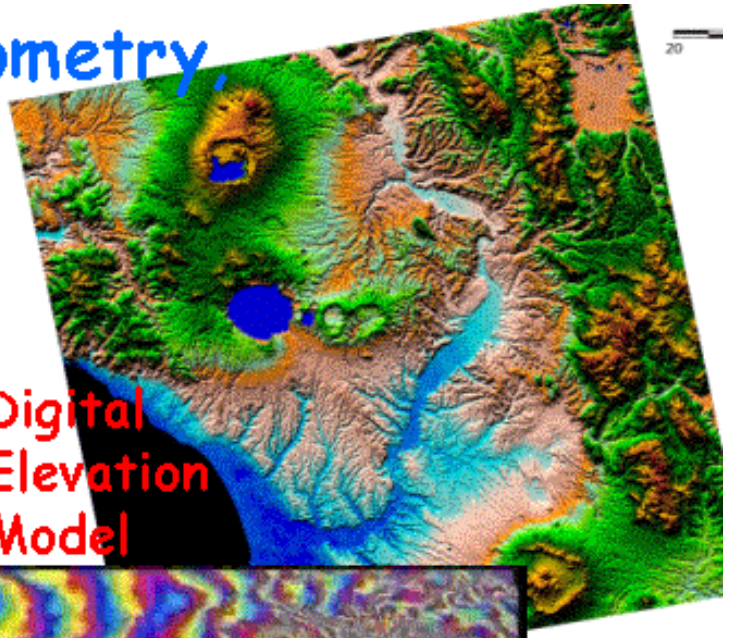
Earth observation applications

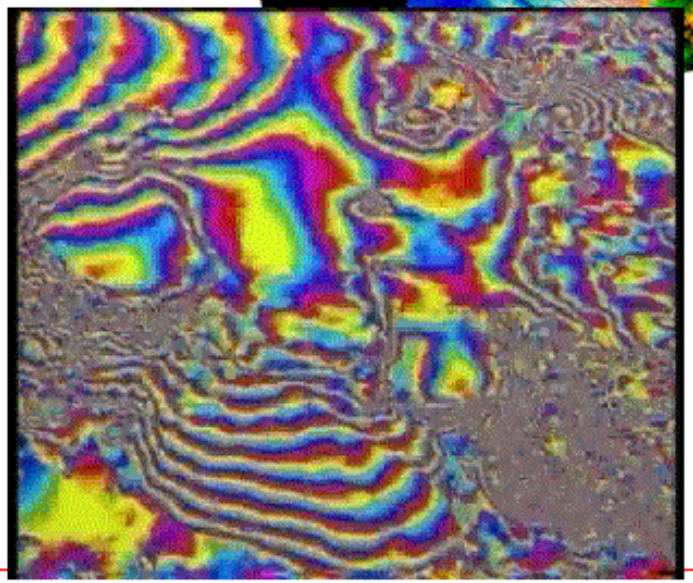
Earth observation applications

Number crunching: interferometry, subsidence, DEM generation

Pomona (Cal): subsidence velocity fields
40 ERS1/2 images (92-99), Ambiguity: 28 mm







GRID requirements:

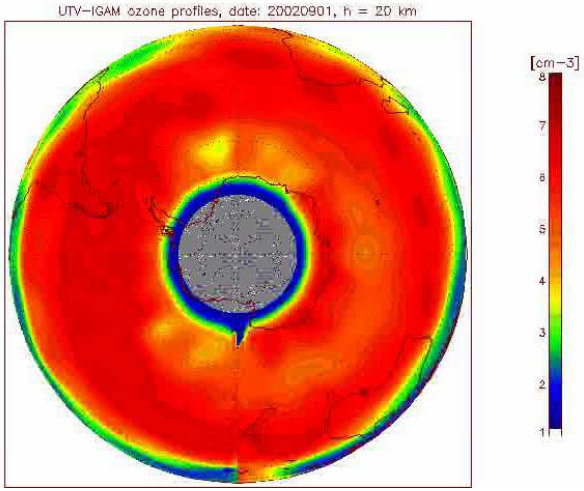
- large data files (10+ GB)
- stages with intensive processing
- science driven value adding

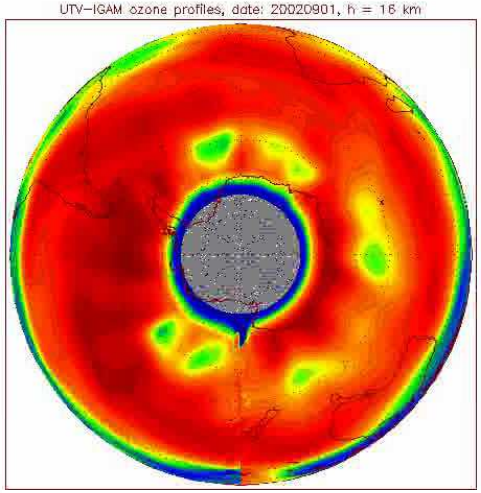
EC Review – 01/03/2002 – WP9 – Earth Observation Applications – n° 5

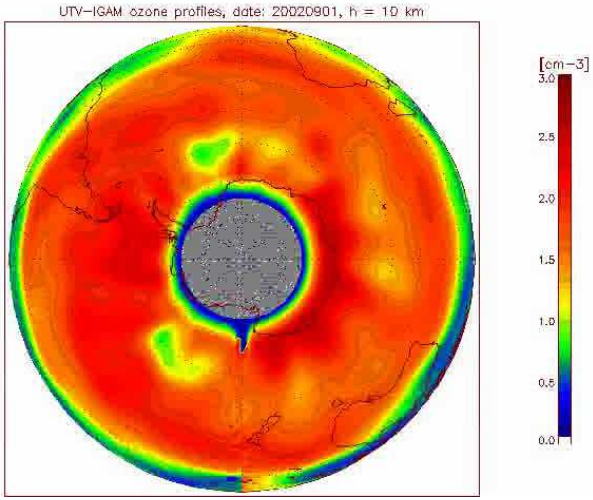
Ozone profiles from GOME (1997-2003)

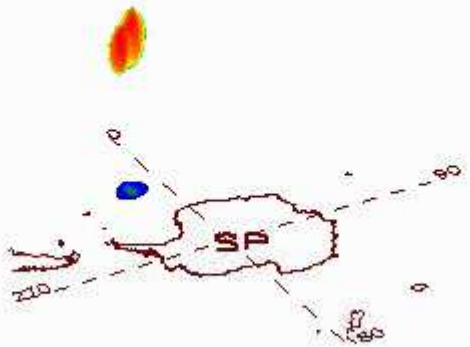
Enabling Grids for E-sciencE

S. Casadio – ESA ESRIN
(GOME 3D Ozone volume over Antarctica - Sept 02, NNO Level 2 products generate in EDG)

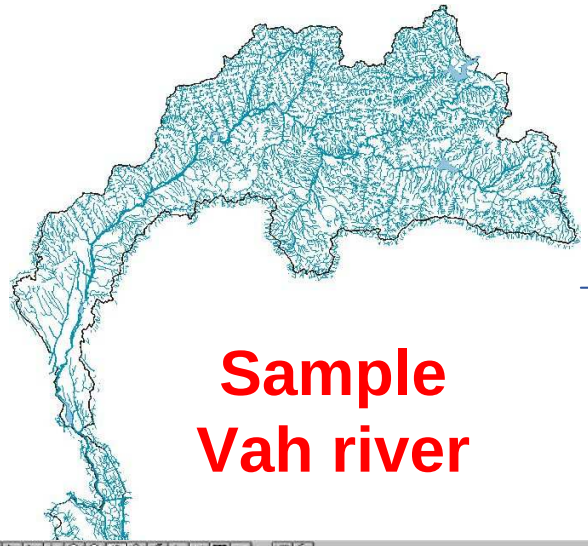




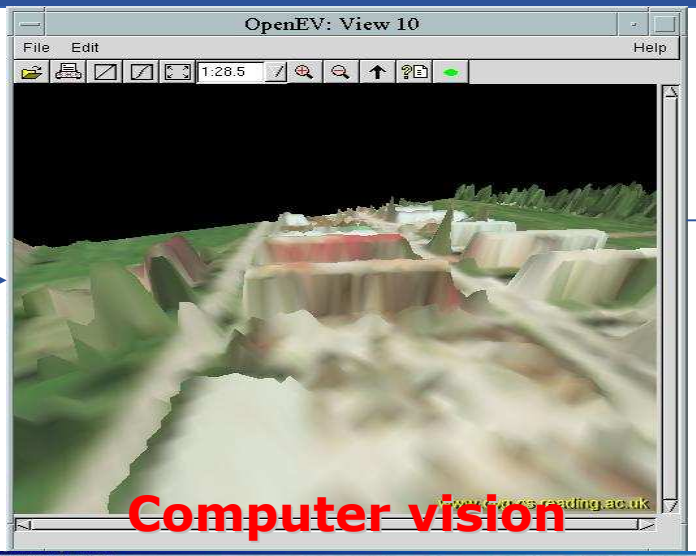




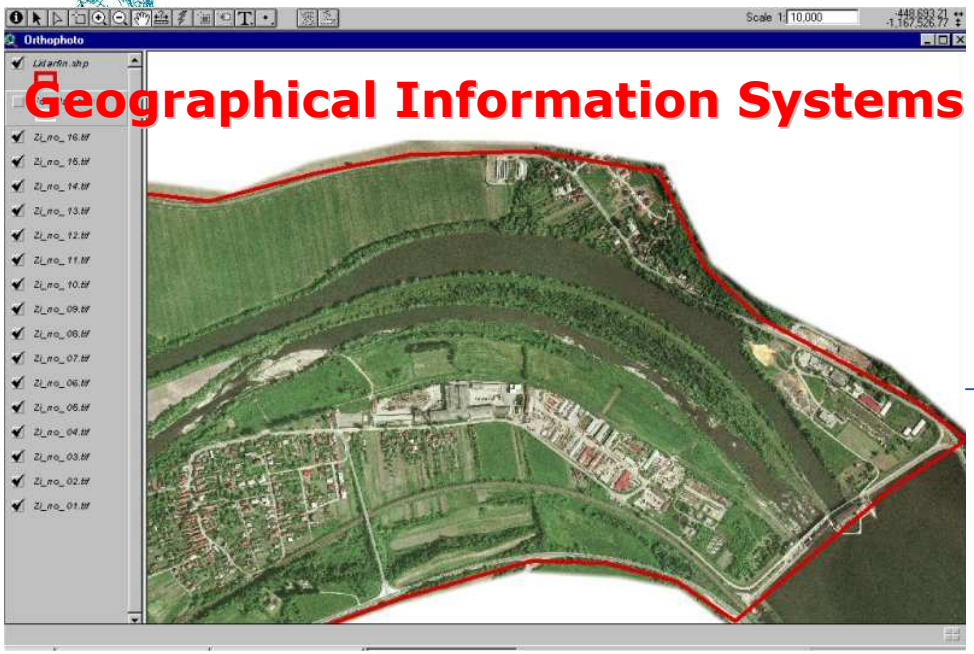
Flood simulation



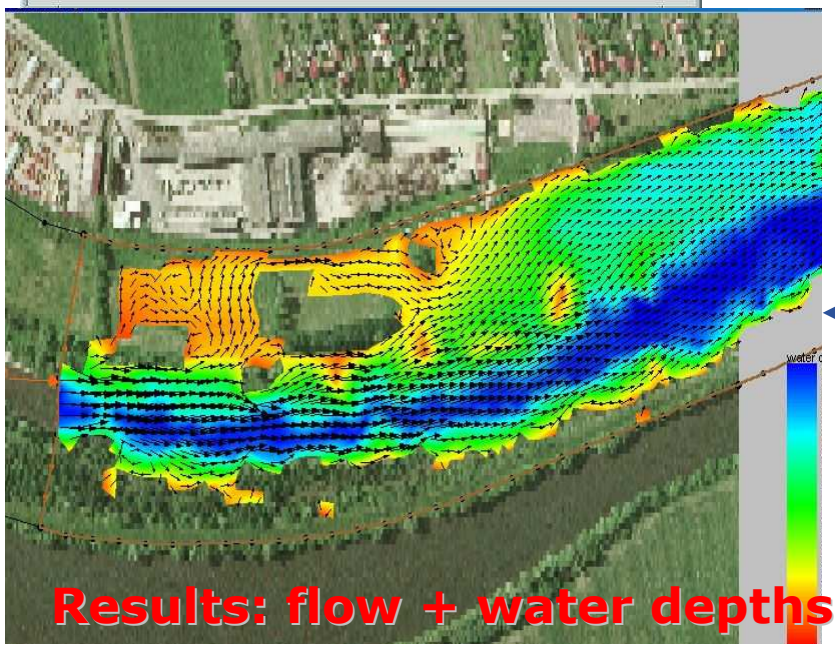
Sample
Vah river



Computer vision

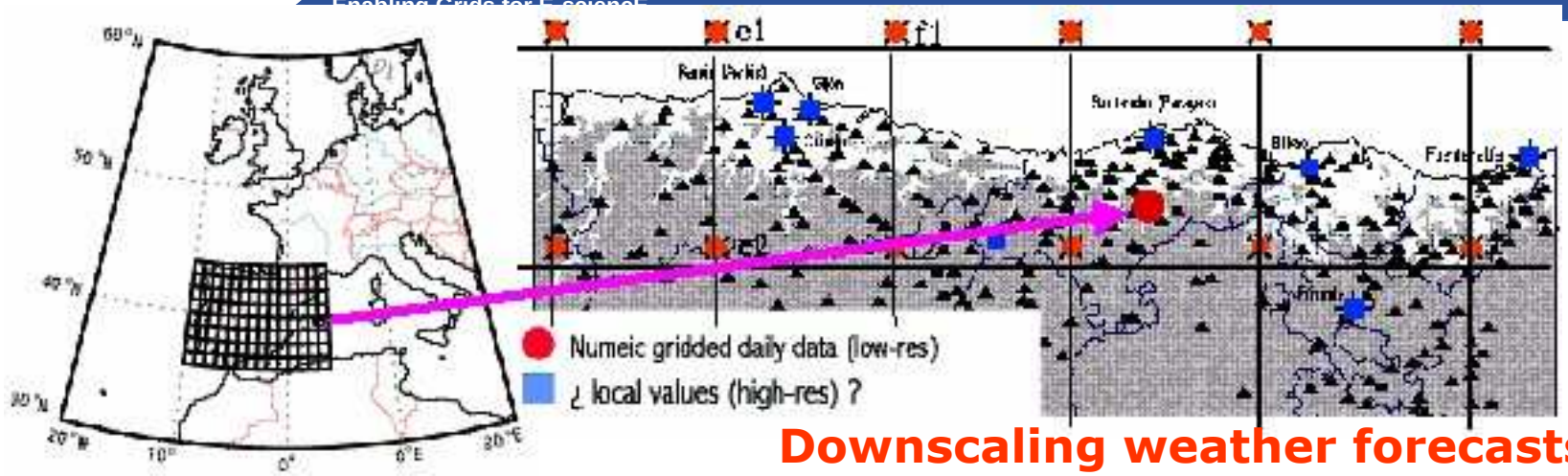


Geographical Information Systems

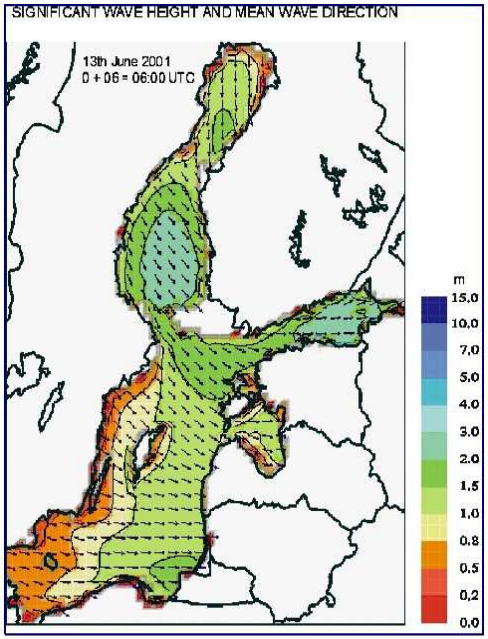


Results: flow + water depths

Weather forecasting

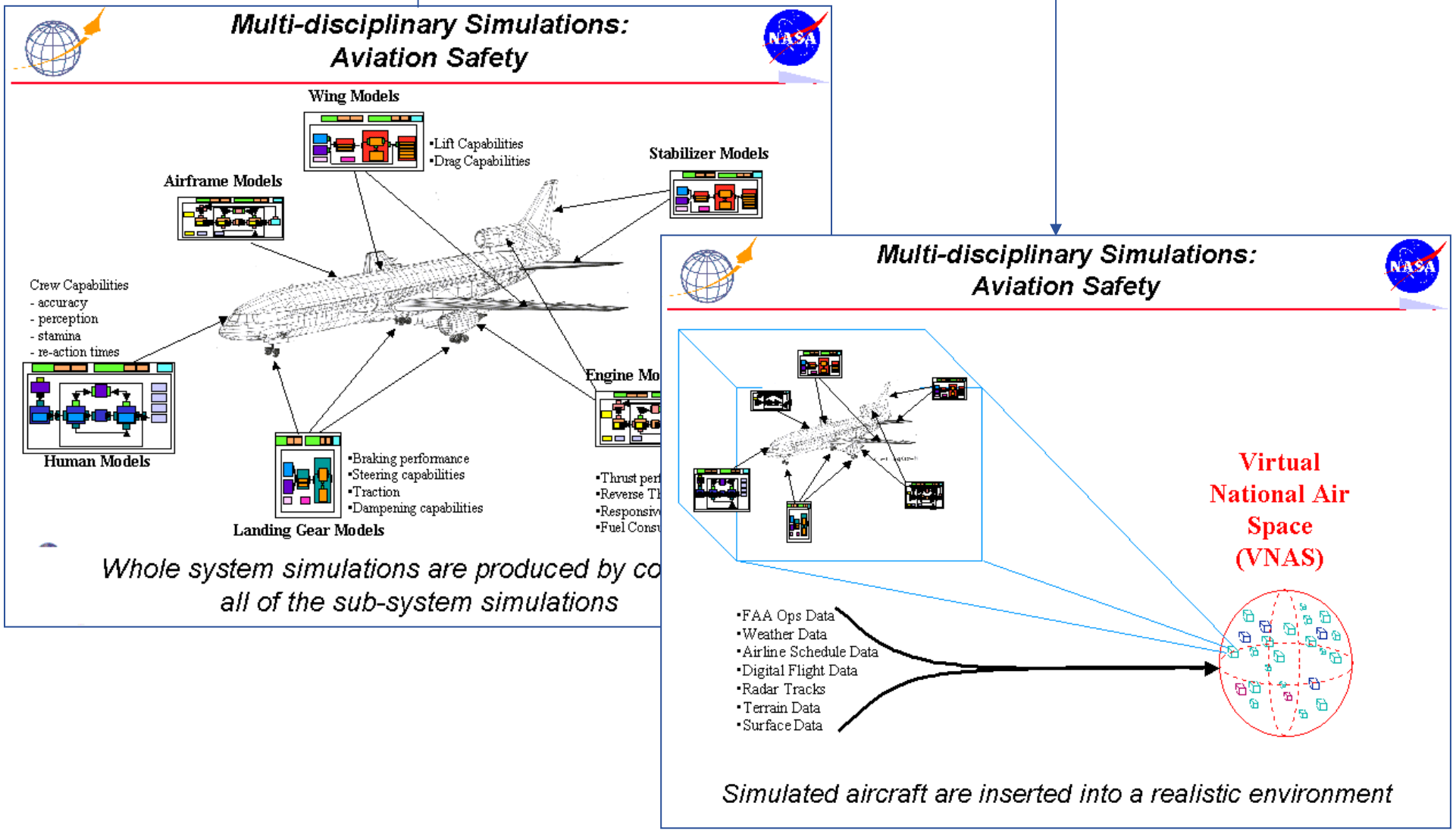


Downscaling weather forecasts



Significant wave height and mean wave direction

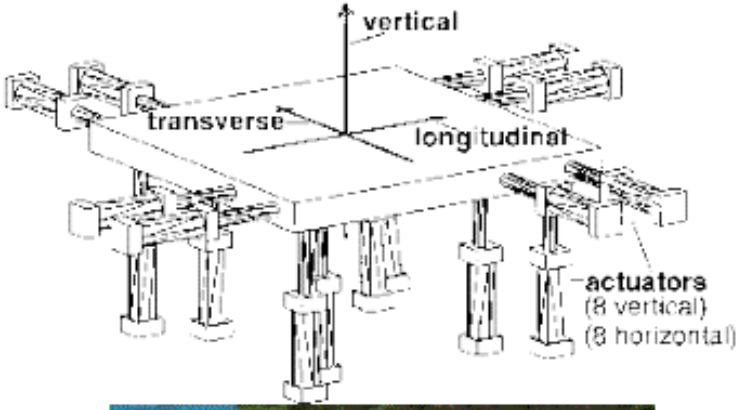
Engineering applications

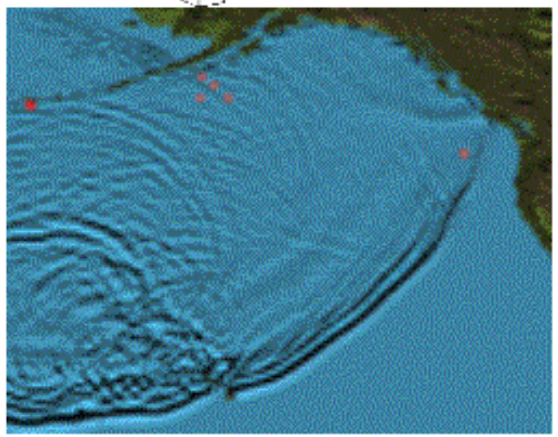


Engineering applications

Network for Earthquake Engineering Simulation

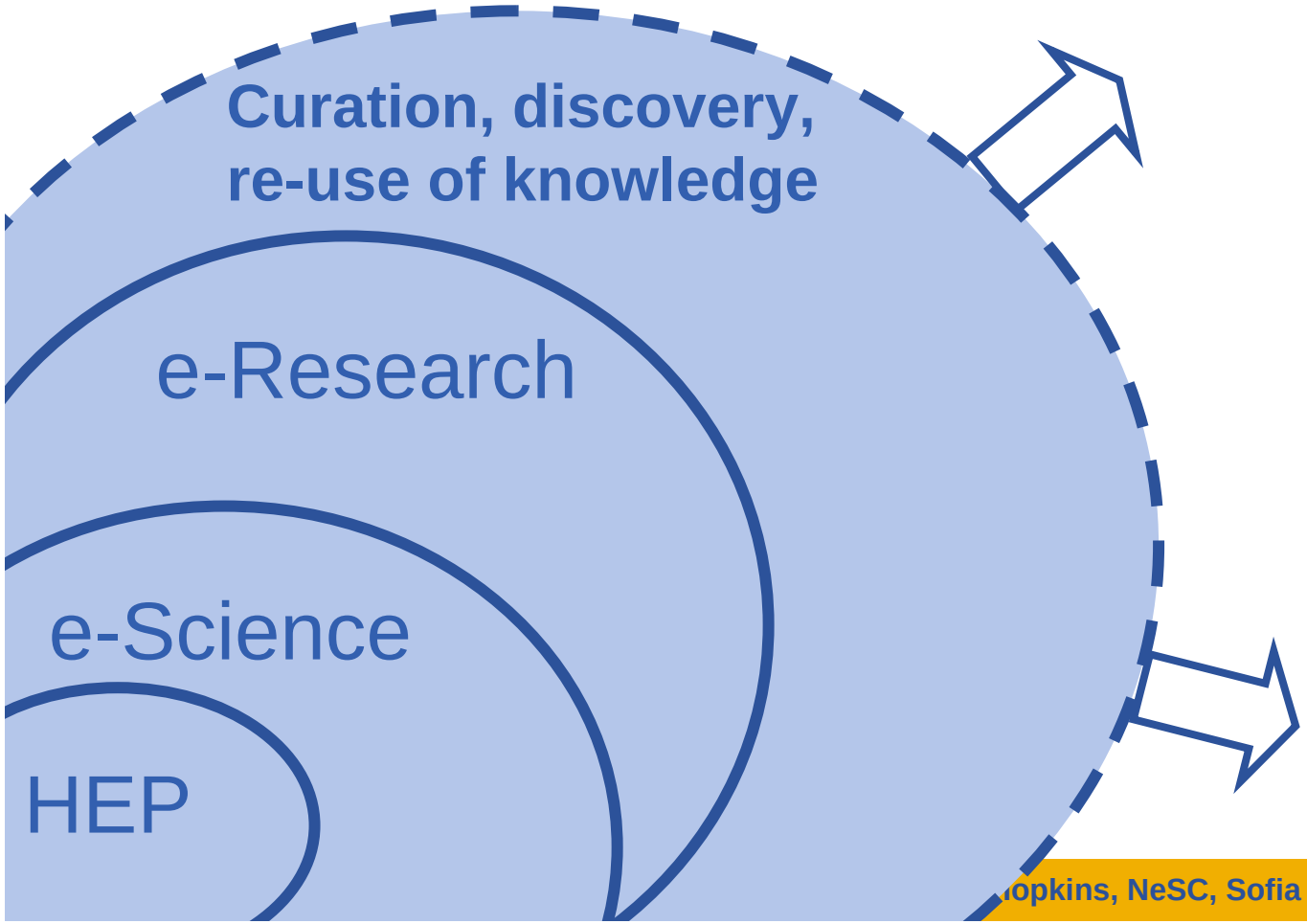
- NEESgrid: national infrastructure to couple earthquake engineers with experimental facilities, databases, computers, & each other
- On-demand access to experiments, data streams, computing, archives, collaboration





NEESgrid: Argonne, Michigan, NCSA, UIUC, USC

The expanding horizons of grids



Grids: will support more than e-Research!

- **Virtual Digital Libraries needed for research as well as learning**
- **Note also: Centrality of curation, preservation**
  - Under-recognised by many researchers
  - Hence the Digital Curation Centre

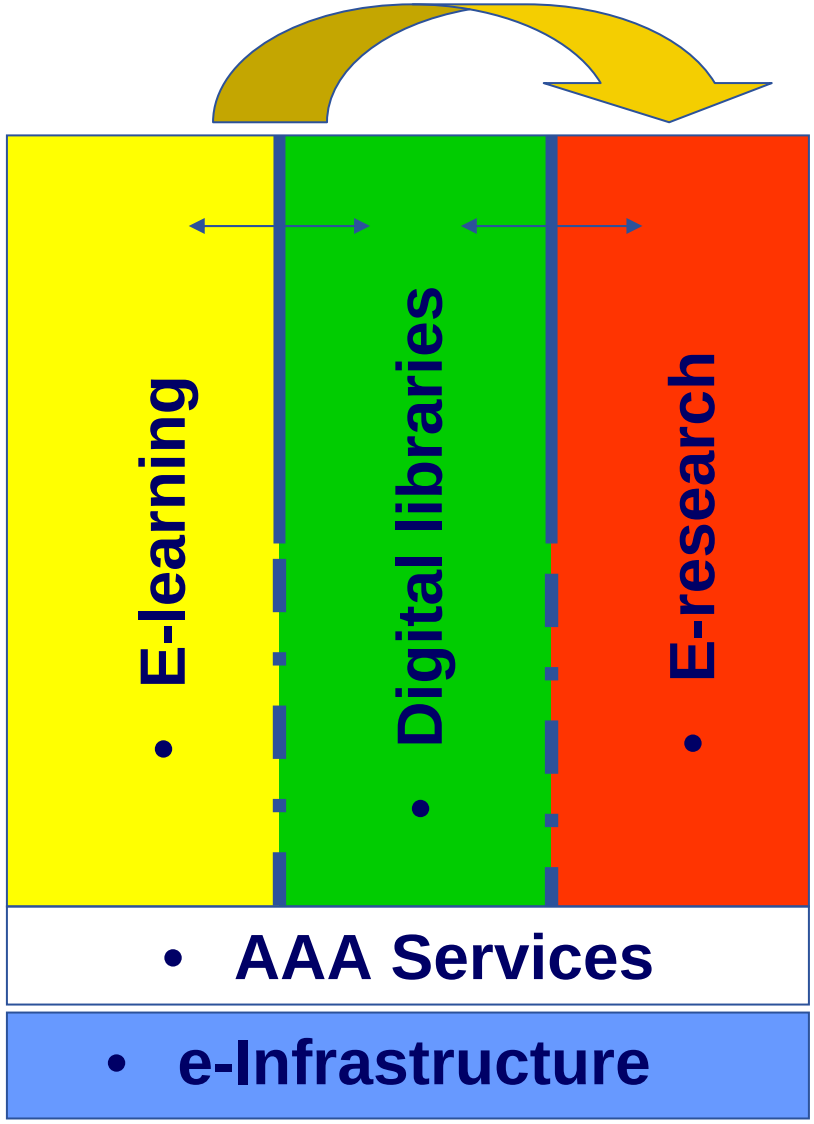


Diagram from a slide by the UK's JISC

Building on e-Infrastructure in 4-D

- **Across geographical distance – networks**
  - Allow remote resources to be accessed
  - SuperJANET, UKLight, GEANT, …
- **Across admin domains – grids**
  - Allow resources in a VO to be shared: virtual computing
- **Across time – data (knowledge) curation**
  - Provides for future research and education
  - Digital Curation Centre (http://www.dcc.ac.uk/)
- **Across disciplines – semantics**
  - How interfaces to services can be understood via a shared ontology, so services can be discovered and used outside their originating community

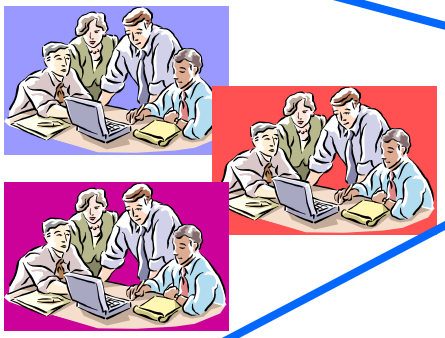
Current “Grid-aware” EU projects for Digital libraries

- **DELOS**
  - Network of excellence exploring technologies for future digital libraries “Future digital libraries should enable any citizen to access human knowledge any time and anywhere, in a friendly, multi-modal, efficient, and effective way”
  - http://www.delos.info/

- **DILIGENT**
  - a DIgital Library Infrastructure on Grid-ENabled Technology that “will allow members of dynamic virtual research organizations to create on-demand transient digital libraries based on shared computing, storage, multimedia, multi-type content and application resources”
  - http://www.diligentproject.org/

DILIGENT DL infrastructure

Consumers



Service A

Service B

DLCreation service

Service C

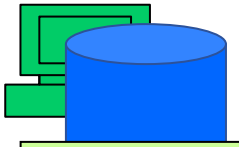
Service D

Service E

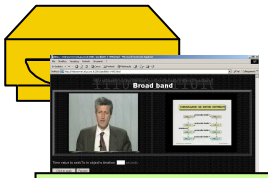
Producers

3D processing

simulation



Feature extraction

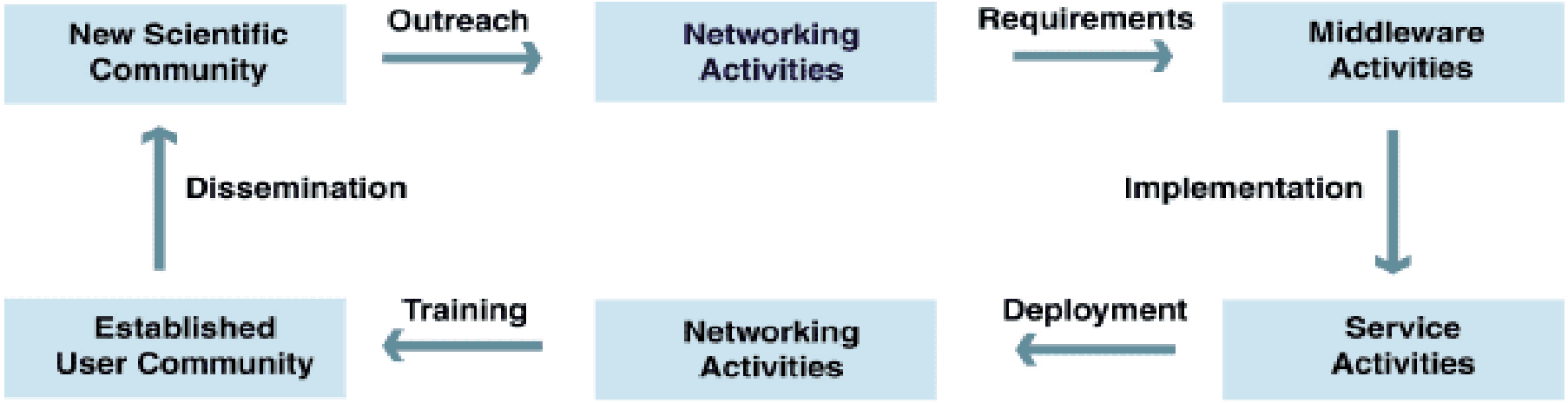


Speech recognition

THE PROCESS FOR NEW VOS

- **Goals - An appreciation of**
  - the range of potential (non-physics) Grid application areas
  - the process by which new application areas are integrated into EGEE as new VOs
- **Outline**
  - Biomed – the other pilot application
  - Some other potential application areas –
    - Earth observation
    - Weather Forecasting
    - Engineering
    - e-Research and beyond
  - **The process for new VO's**
  - The up-coming VOs –
    - Computational Chemistry
    - Earth Science
    - Astrophysics

Identification and integration of new communities: EGEE virtuous cycle



- Virtuous cycle concept is described in the project Technical Annex
- It describes the role of the different project activities to help new communities to successfully deploy applications on EGEE infrastructure
- As the first open multidisciplinary e-infrastructure in the world, EGEE has to invent the implementation of the virtuous cycle

New communities identification

- **Through training, dissemination and outreach, communities already using advanced computing and keen to use EGEE infrastructure are identified**
- **These communities are encouraged to prepare a document describing their interest to use EGEE**
- **A scientific advisory panel (EGAAP) assesses and chooses among the interested communities the ones which seem the most mature to deploy their applications on EGEE**

EGAAP

- **EGEE Generic Applications Advisory Panel is the entry door for new applications that want to be deployed on the EGEE infrastructure**

- **Important step in the EGEE virtuous cycle**
  - Encourages communities to submit a well documented proposal
  - Fosters discussion on the added value brought by the Grid to the applications
  - Points out needs and resources for migration and deployment for each application
  - Prioritizes the deployment of the selected applications
  - Monitors the progress of the selected portfolio

- **Participation in EGAAP of 5 external members is useful to reach out to new communities**

EGEE Industry Forum

- **Objectives:**
  - To promote and disseminate Grid concepts towards industry and service groups
  - To raise the awareness of EGEE within industry
  - To encourage businesses to participate in the project

- **Members:** **interested companies having activities in Europe**

- **Activities:**
  - Organisation of a meeting twice a year
  - Quarterly newsletter
  - Participation to EGEE working groups (EGAAP, Project Technical Forum, EGEE Phase 2, Security group)
  - Internal Working groups
    - Technical aspects of Grid
    - Business models and economical aspects

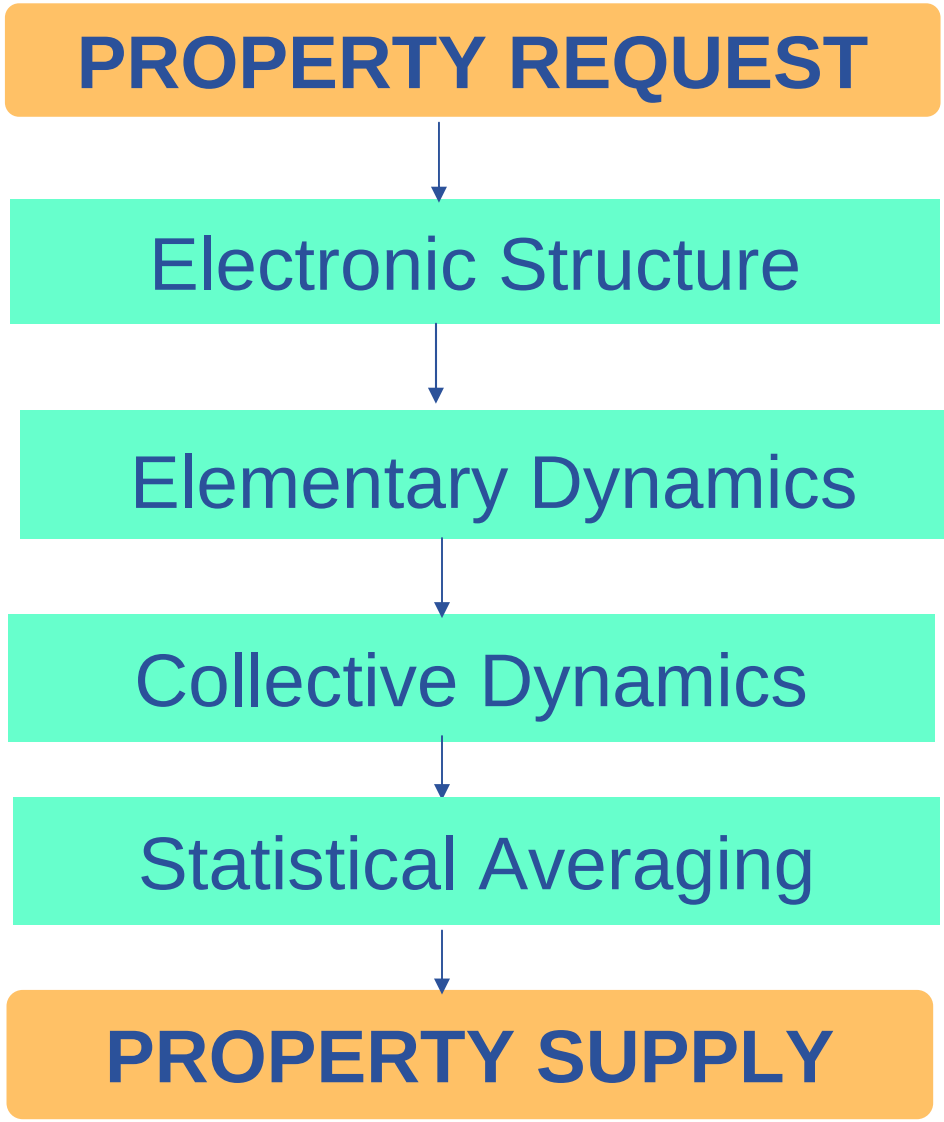
Up-COMING VOs

- **Goals - An appreciation of**
  - the range of potential (non-physics) Grid application areas
  - the process by which new application areas are integrated into EGEE as new VOs
- **Outline**
  - Biomed – the other pilot application
  - Some other potential application areas –
    - Earth observation
    - Weather Forecasting
    - Engineering
    - Art
  - The process for new VO's
  - **The up-coming VOs –**
    - **Computational Chemistry**
    - **Earth Science**
    - **Astrophysics**

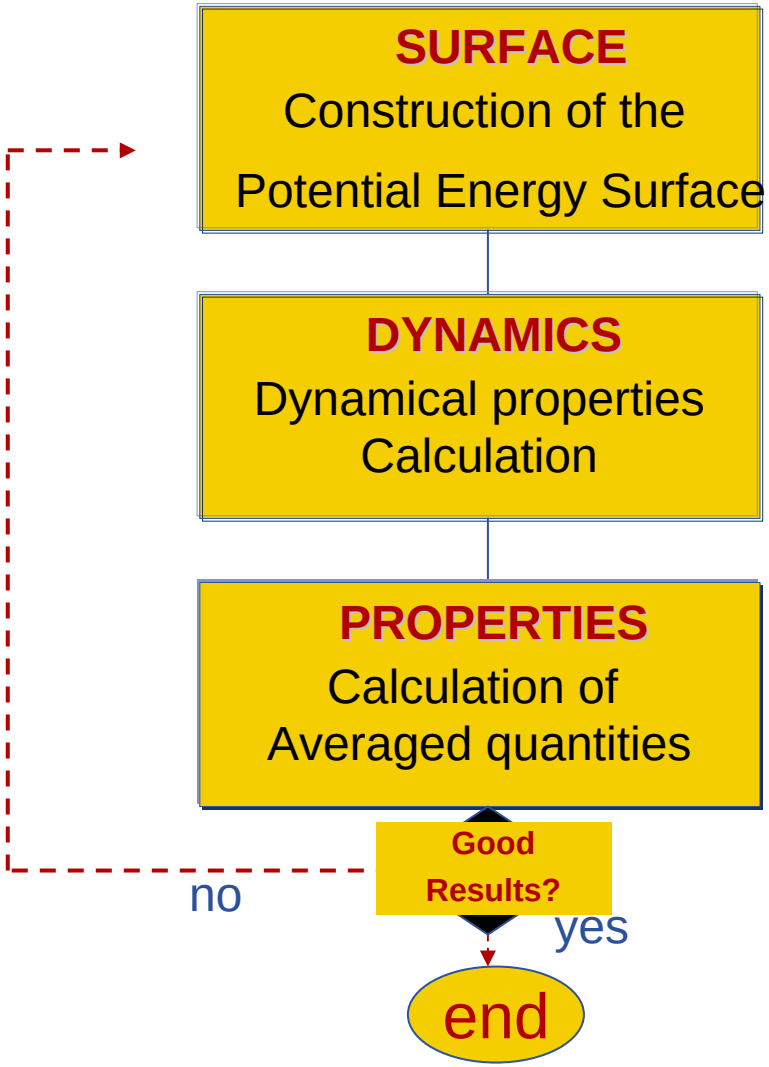
Computational Chemistry

GEMS,
Grid Enabled Molecular Simulations

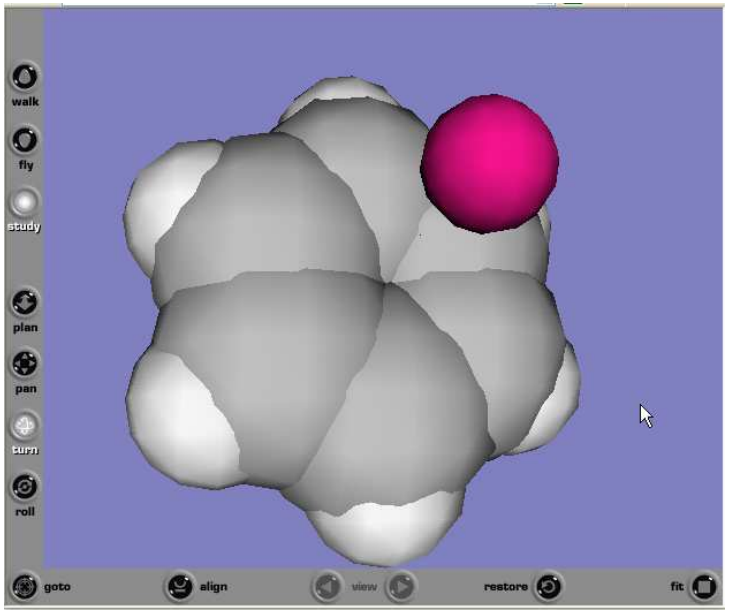
The Molecular Simulator



Computational Chemistry: molecular simulator



Ar - Benzene



The mathematical formalism

$$i\eta\frac{\partial}{\partial t}\Psi(\{W\},\{w\},t)=\hat{H}\Psi(\{W\},\{w\},t)$$

Separation of electronic and nuclear motions

Electronic Schrödinger equation:

$$\hat{H}_{elec}\Psi_n(\{w\};\{W\})=E_n(\{W\})\Psi_n(\{w\};\{W\})$$

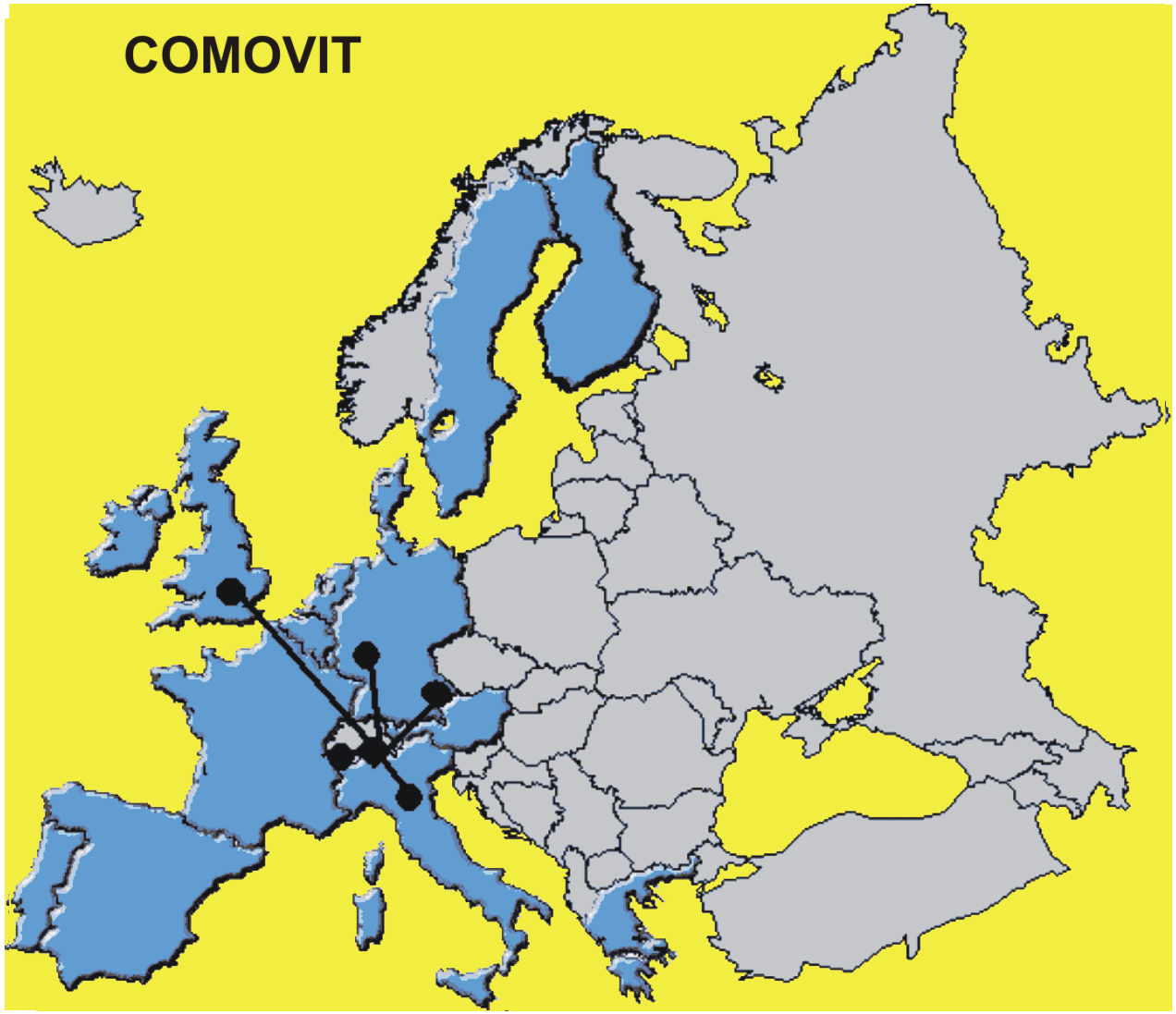
Nuclear Schrödinger equation:

$$\hat{H}_n\chi_n(\{W\},t)=i\eta\frac{\partial}{\partial t}\chi_n(\{W\},t)$$

+ Statistical averaging for beam conditions

The CHEMISTRY community

- Simbex
- Murqm
- Dirac
- Elchem
- Icab
- Dysts
- Comovit

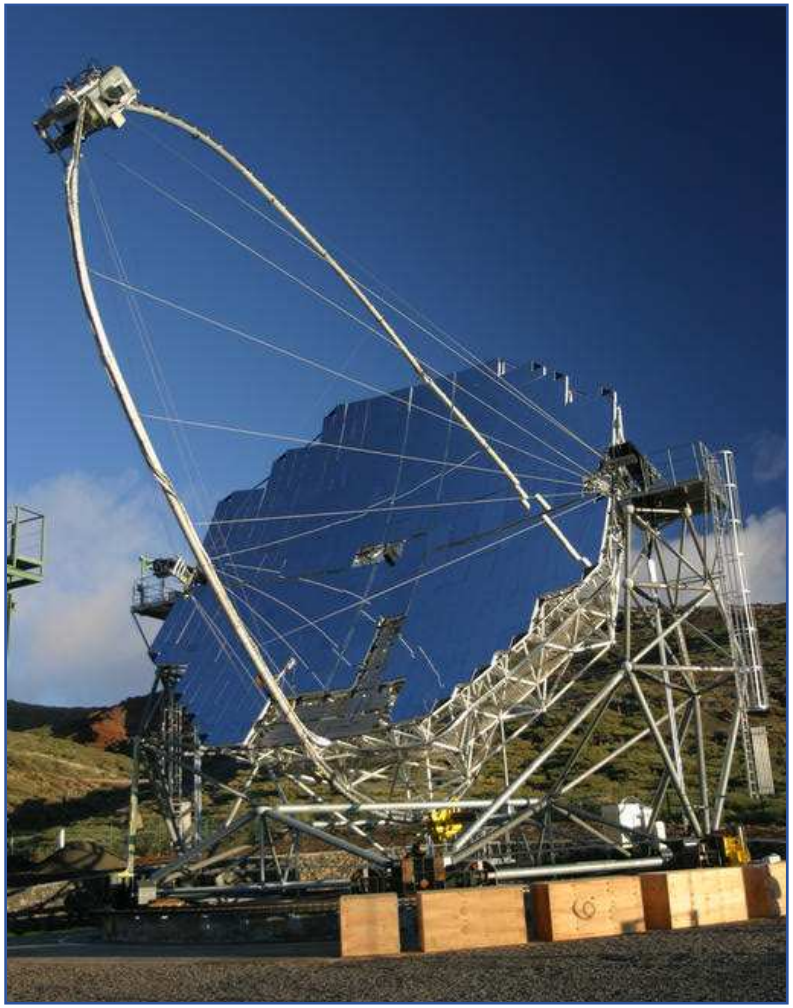


New projects and laboratories

- 3 Computer Centres
- New electronic structure programs (MOLCAS, DIRAC, DALTON, COLUMBUS, MR-CCSD).
- New Dynamics programs (AMD, TPS, KMC, condensed phase).
- Chemical knowledge semantic web (molecular structures, apparatuses, processes).

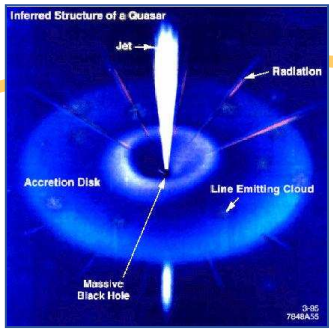
ASTROPHYSICS
The MAGIC telescope

- Largest Imaging Air Cherenkov Telescope
  (17 m mirror dish)
- **Located on Canary Island**
  La Palma **(@ 2200 m asl)**
- **Lowest** energy threshold **ever obtained with a Cherenkov telescope**
- **Aim: detect** γ**-ray sources in the unexplored energy range:**
  30 *(10)*-> 300 GeV

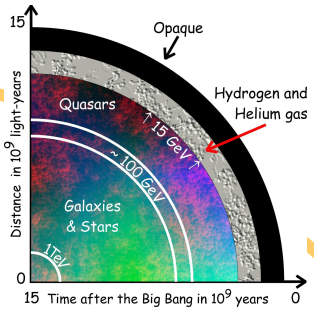


The MAGIC Physics Program

- Pulsars



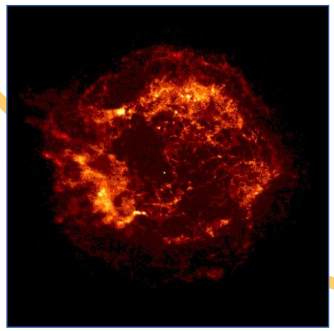
- AGNs



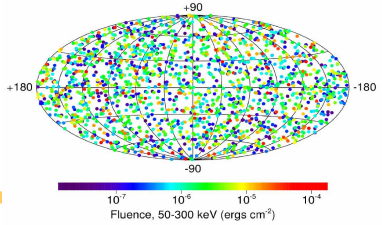
- Cosmological γ-Ray Horizon

- Origin of Cosmic Rays

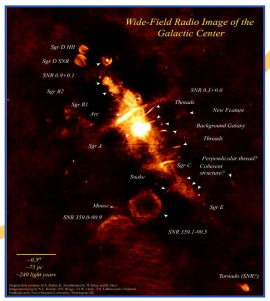
- Tests of Quantum Gravity effects



- SNRs



- GRBs



- Cold Dark Matter

Computational Chemistry Achievements & Issues

Achievements

- **Cluster of 13 nodes + CE + SE + VOMS server has been deployed in GILDA for dedicated use by CompChem.**
- **Grid based Molecular Simulator (GEMS) ported onto the GILDA test cluster and interfaced to GENIUS**
- **The CompChem VO has been activated**
- **Work in hand now to move to production service**

Issues

- **Requirements for interactive work**
  - Outbound connectivity of worker nodes
  - Fast turnaround in jobs
- **Access to licensed software**

Earth Science Achievements & Issues

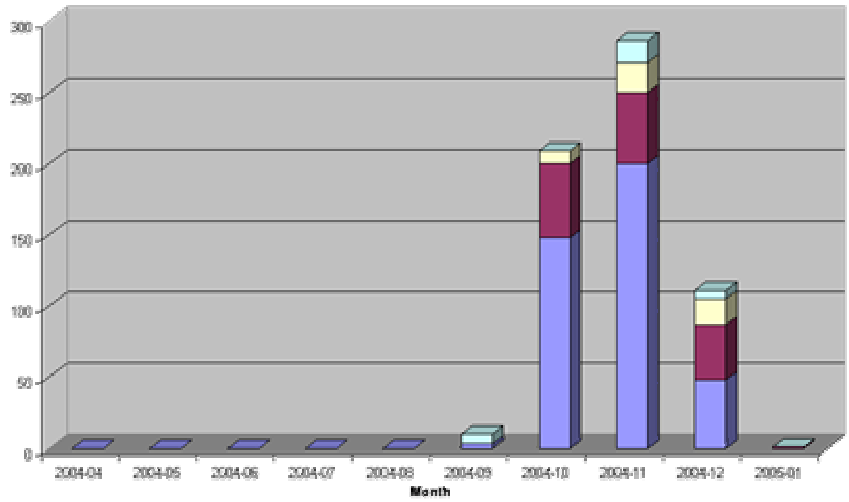
Achievements

- ESR (Earth Sciences Research) VO at SARA created in July 2004 and is functional using EGEE resources
  - 17 registered users from 6 countries
- The EGEODE (Expanding GEOsciences on DEmand) VO created at IN2P3 (Lyon) in mid-October for CGG and Geocluster partners
  - Preparation to migration to EGEE Production Service
- Important EGEODE application deployed on GILDA and demonstrated at the 2nd EGEE Conference in The Hague using the GENIUS portal
- Production of ozone profiles from the satellite experiment GOME and their validation by using LIDAR data run on EGEE production service

Issues

- Need secure access to data and metadata for authorised groups/sub-groups
- Access to licensed software

Number of jobs submitted by ESR VO members



MAGIC Achievements & Issues

Achievements

- **A Magic Virtual Organisation already exists in EGEE**
  - VO server is hosted by SARA/NIKHEF
  - Successful first running in GILDA as well as in Crossgrid testbed using LCG-2 middleware
- **Developments underway for EGEE data challenge in early 2005**
  - CNAF will support the Magic VO with a Resource Broker
  - PIC will support the Magic VO with storage and the RLS
  - CNAF, PIC and GridKA will provide CPU
  - GILDA can be used for the first test as well

Issues

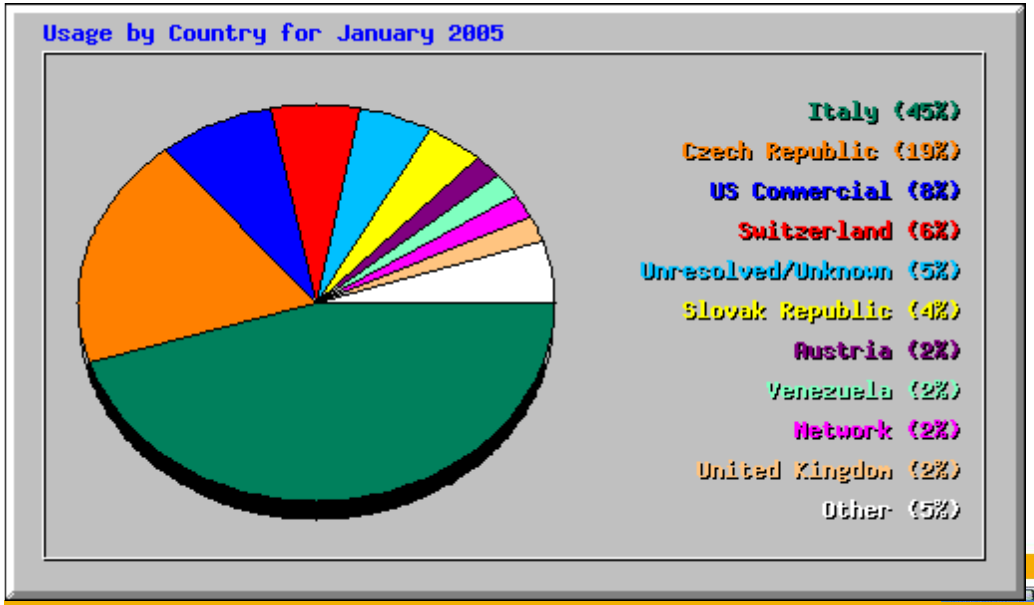
- **Education**
  - 'EGEE for dummies'
- **Getting extra EGEE resources for data challenge**
  - Precise 'process' definition and its execution

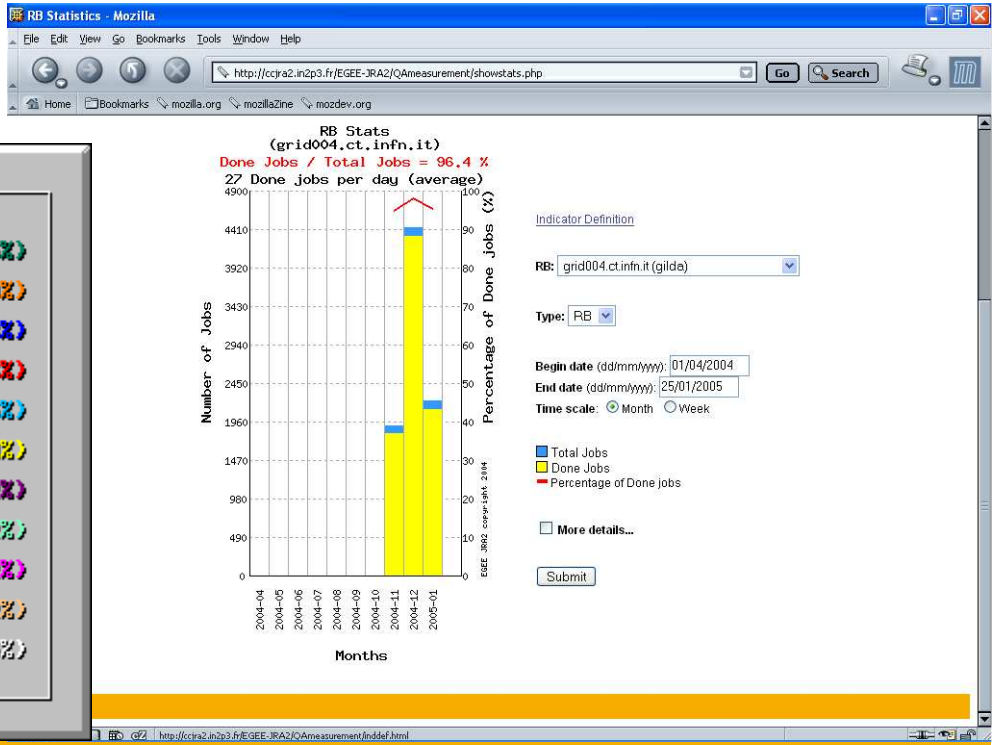
GILDA, an infrastructure for dissemination and demonstration

- **Goals**
  - Demonstration of grid operation for tutorials and outreach
  - Initial deployment of new applications for testing purposes
- **Key features**
  - Initiative of the INFN Grid Project using LCG-2 middleware
  - On request, anyone can quickly receive a grid certificate and a VO membership allowing them to use the infrastructure for 2 weeks
  - Certificate expires after two weeks but can be renewed
  - Use of friendly interface: Genius grid portal
- **Very important for the first steps of new user communities on to the grid infrastructure**

GILDA numbers

- 14 sites in 2 continents
- >1200 certificates issued, 10% renewed at least once
- >35 tutorials and demos performed in 10 months
- >25 jobs/day on the average
- Job success rate above 96%
- >320,000 hits on the web site from 10's of different countries
- >200 copies of the UI live CD distributed in the world





NA4 Applications and GILDA

- **7 Virtual Organizations supported:**
  - Biomed
  - Earth Sciences
    - Earth Science Academy (ESR)
    - Earth Science Industry (CGG)
  - Astrophysics
    - Astroparticle Physics (MAGIC)
    - Astrophysics (PLANCK)
  - Computational Chemistry (GEMS)
  - Grid Search Engines (GRACE)

- **Development of complete interfaces with GENIUS for 3 Biomed Applications: GATE, hadronTherapy, and Friction/Arlecore**

- **Development of complete interfaces with GENIUS for 4 Generic Applications: EGEODE (CGG), MAGIC, GEMS, and CODESA-3D (ESR) (see demos!)**

- **Development of complete interfaces with GENIUS for 16 demonstrative applications available on the GILDA Grid Demonstrator (https://grid-demo.ct.infn.it)**

Summary

- EGEE and grids – not just physics
- For communities to benefit they need to know what grids can do for them – dissemination
- Many communities are beginning to adopt the grid
- EGEE has a mechanism for assisting communities onto the grid

The End

- **The end**