



Enabling Grids for E-science

EGEE Applications (some success stories)

Roberto Barbera

University of Catania and INFN

ISSGC05

Vico Equense, 20.07.2005

www.eu-egee.org



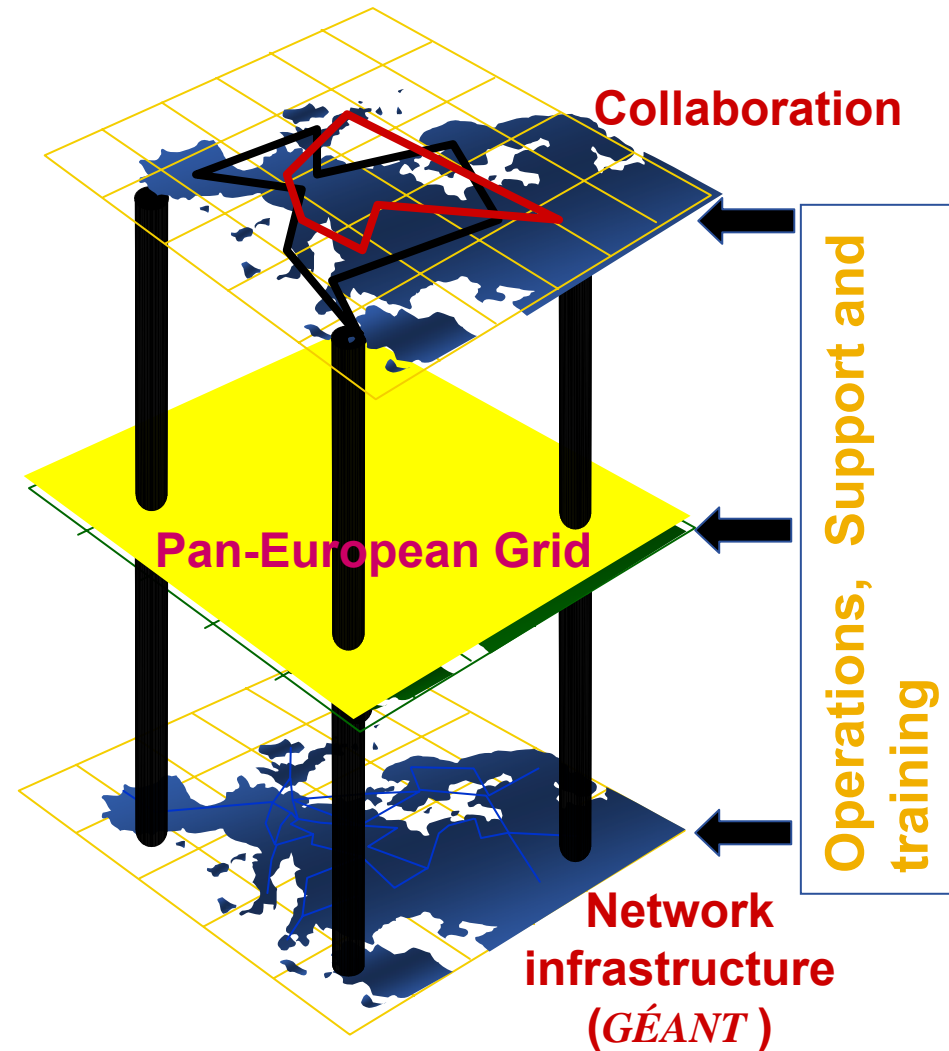
Information Society



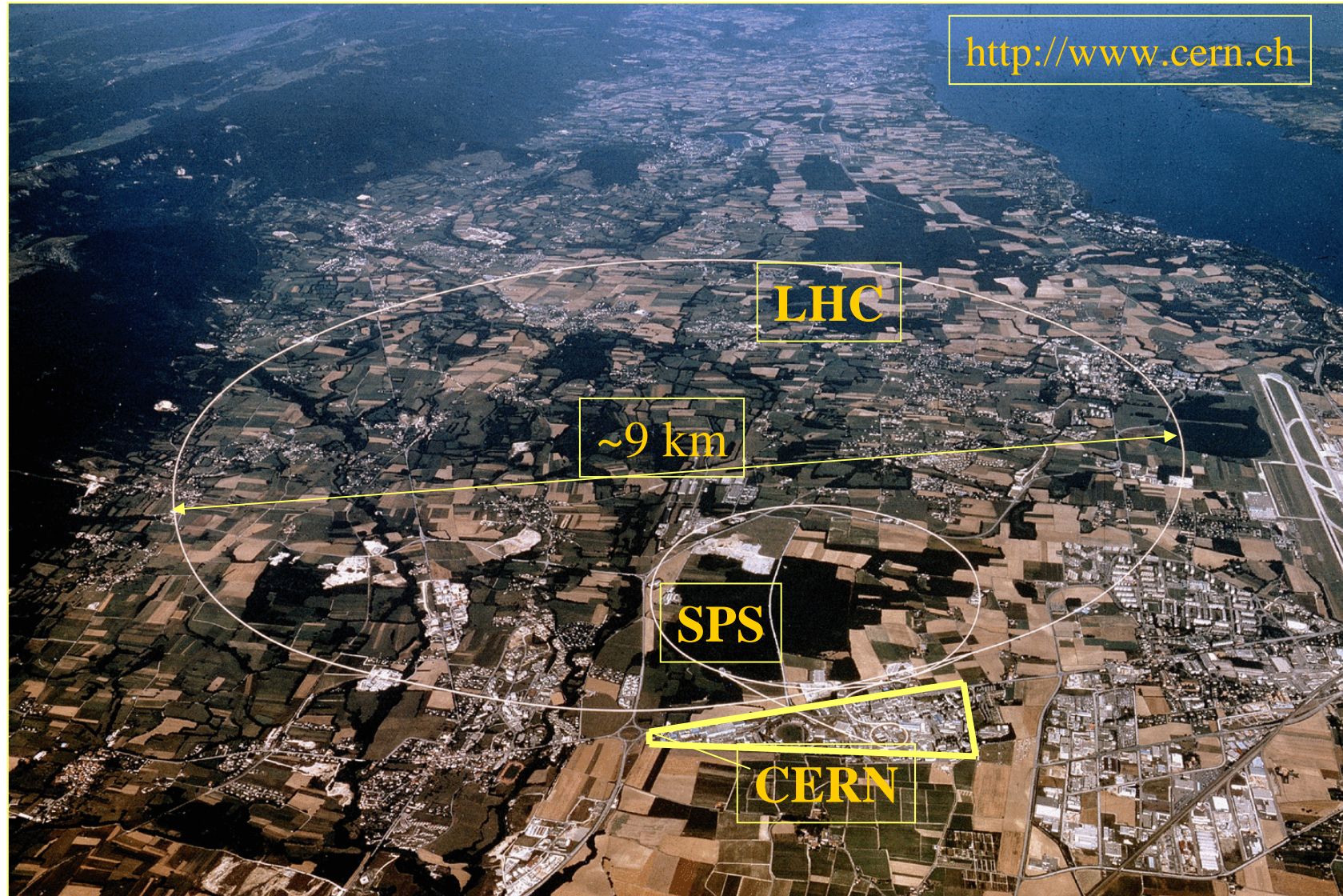
- **Introduction**
- **EGEE pilot applications: HEP and Biomed**
- **EGAAP and EGEE “generic applications”**
- **Summary**

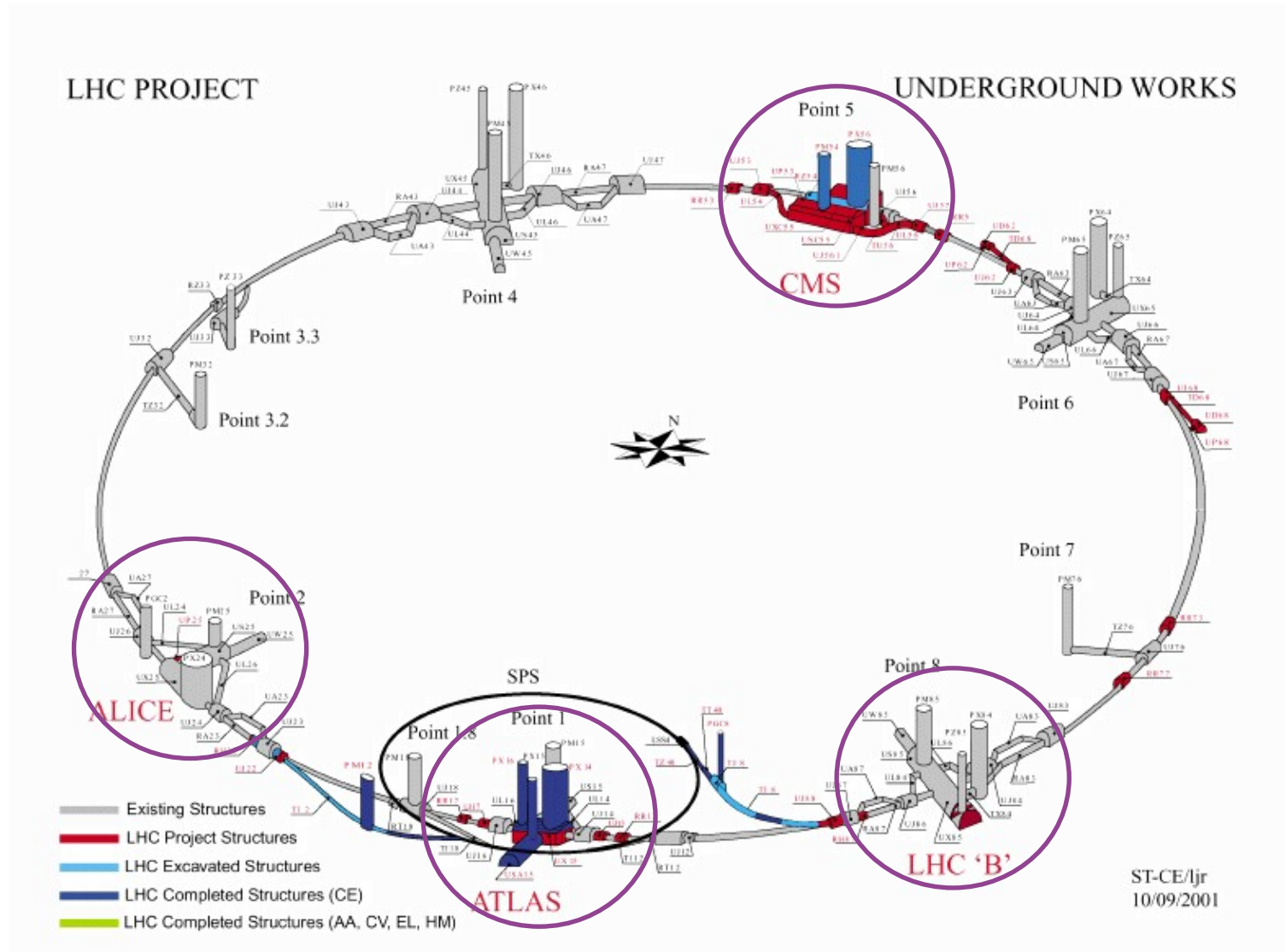
Build a large-scale production grid service to:

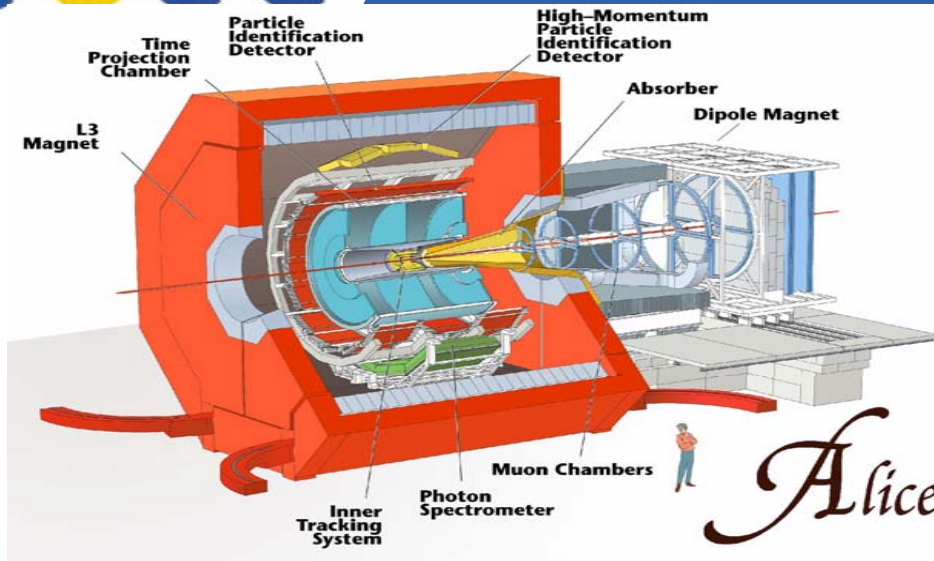
- Underpin European science and technology
- Link with and build on national, regional and international initiatives
- Foster international cooperation both in the creation and the use of the e-infrastructure



- **Three application groups**
 - High Energy Physics pilots
 - Biomedical application pilots
 - Generic applications (catch-all)
- **Multiple infrastructures, two middlewares**
 - EGEE LCG2 production infrastructure
 - GILDA LCG2/gLite integration infrastructure
 - gLite testbeds (development/testing/certification)
- **Many users**
 - broad range of needs
 - different communities with different background and internal organization







Total weight 10.000t
External diameter 16,00m
Total length 25m
Magnetic Field 0.2-0-5Tesla

ALICE On-line System
multi-level trigger to filter background and reduce the amount of data

8 kHz (160 GB/sec)

level 0 – custom hardware

200 Hz (4 GB/sec)

level 1 – embedded processors

30 Hz (2.5 GB/sec)

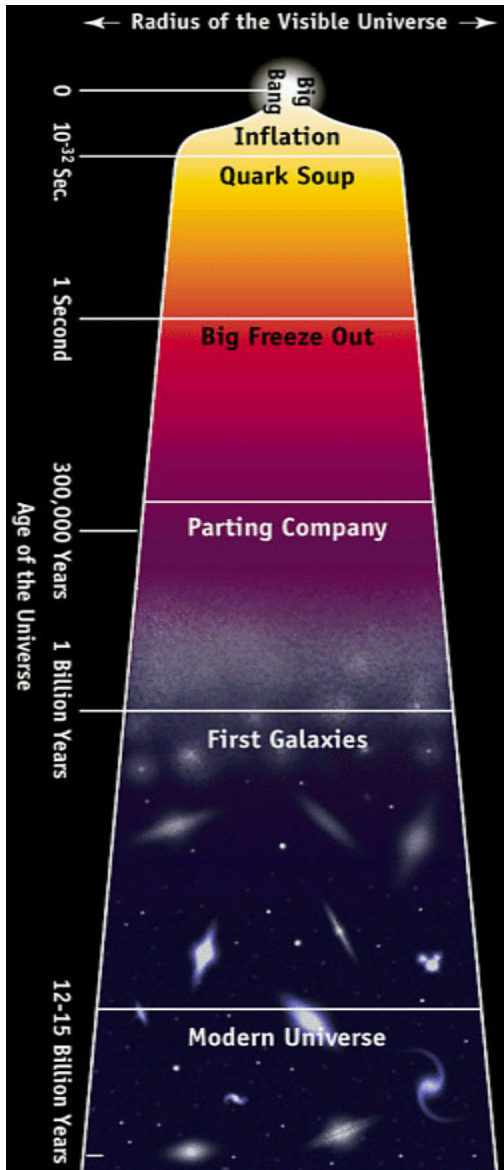
level 2 – PC's

30 Hz
(1.25 GB/sec)

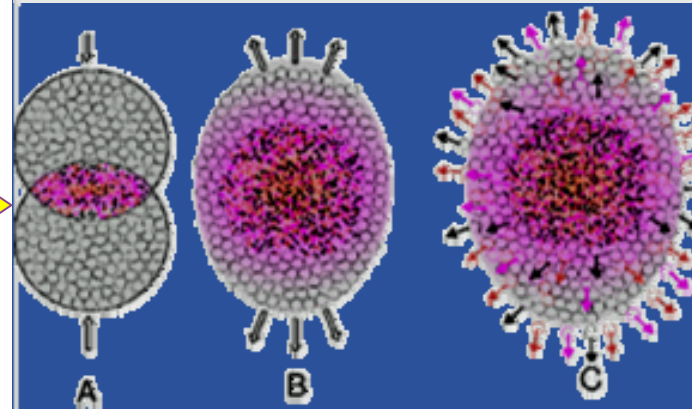
data storage & off-line analysis

ALICE Collaboration already includes 1223 peoples from 85 institutes in 27 countries

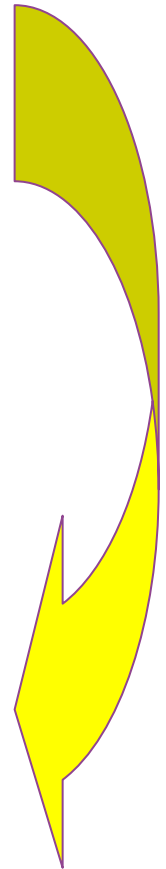
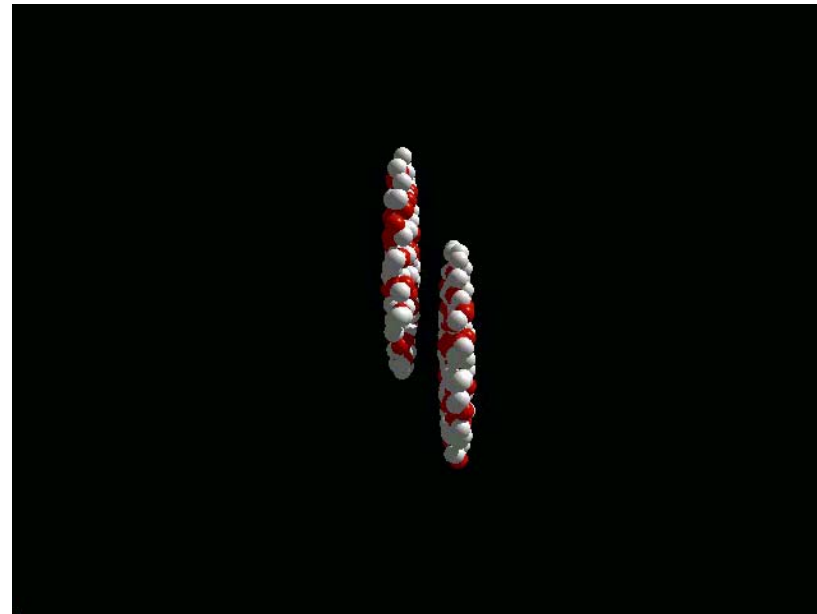
The Big Bang



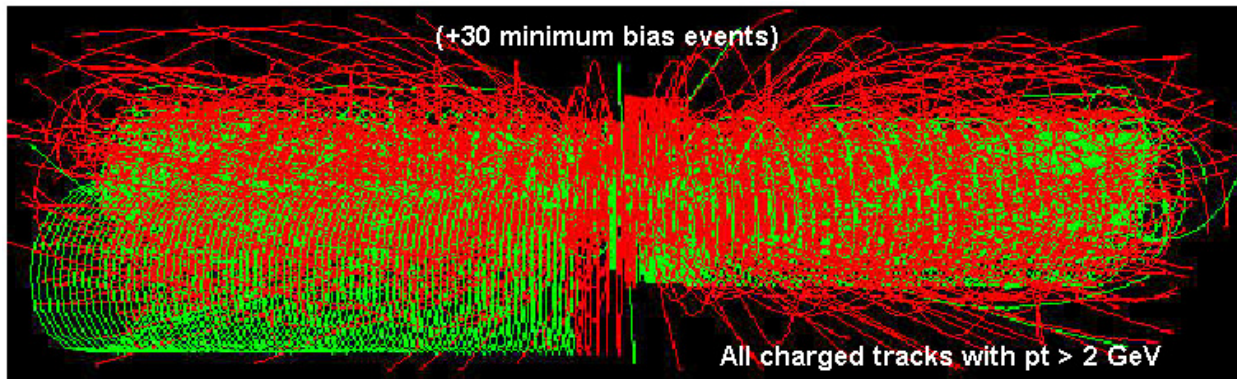
The Little Bang



Pb+Pb @ LHC (5.5 A TeV)



Starting from this event...

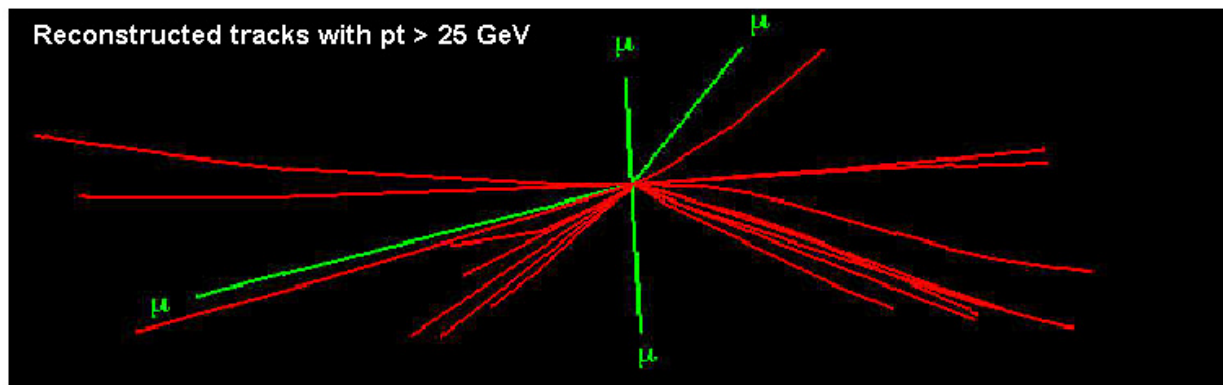


Selectivity: 1 in 10^{13}

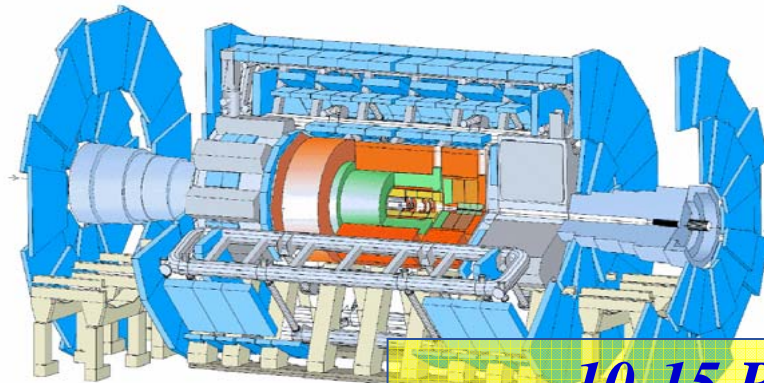
Like looking for 1 person in a thousand world populations!

Or for a needle in 20 million haystacks!

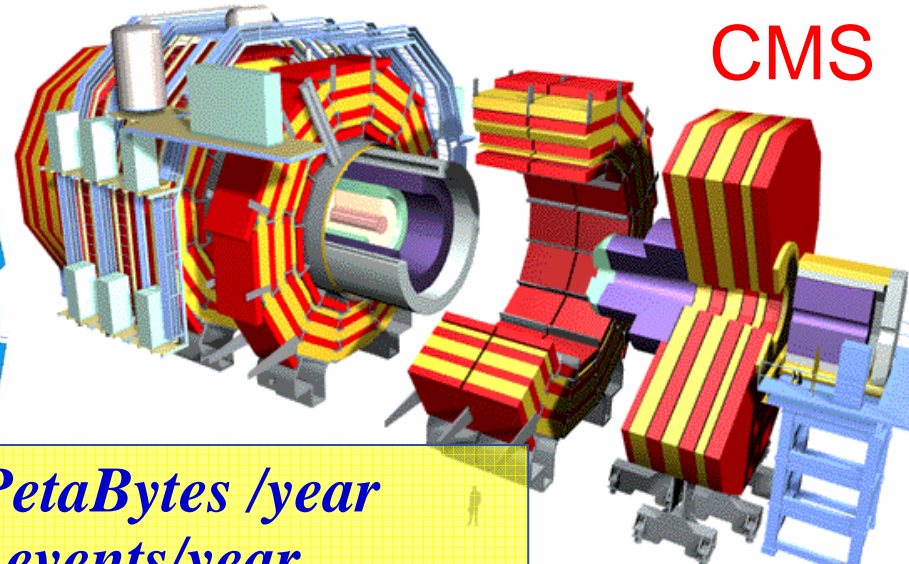
You are looking for this “signature”



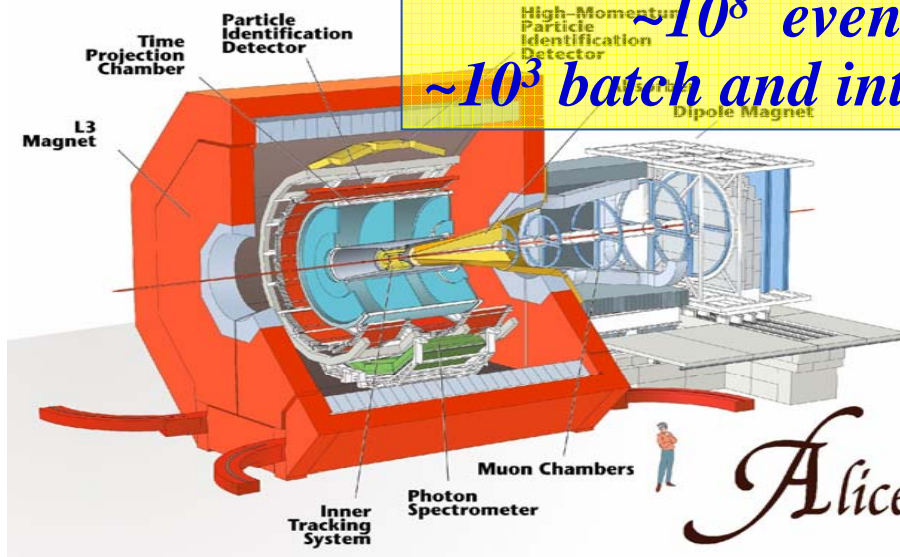
ATLAS



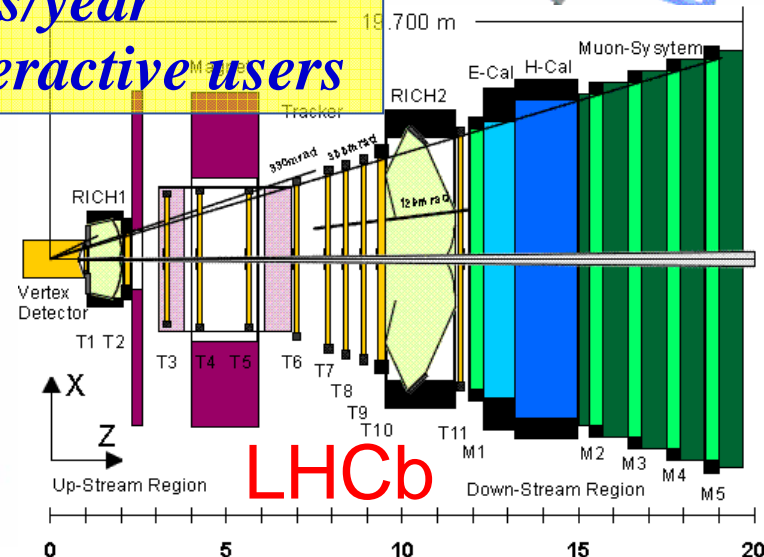
CMS



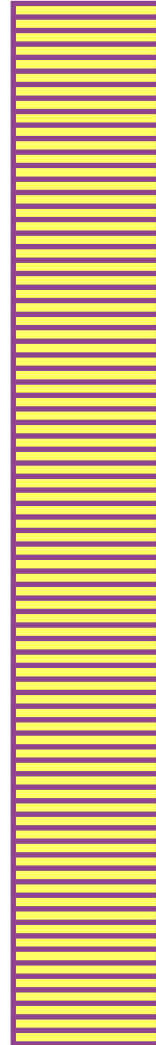
~10-15 PetaBytes /year
~10⁸ events/year
~10³ batch and interactive users



Alice



**10-15 Petabytes
~20.000.000 CD-
ROM**



**10 times the
Eiffel Tower
~3000 m**

The LHC “web” (2002 data)



Europe: 267 institutes, 4603 users

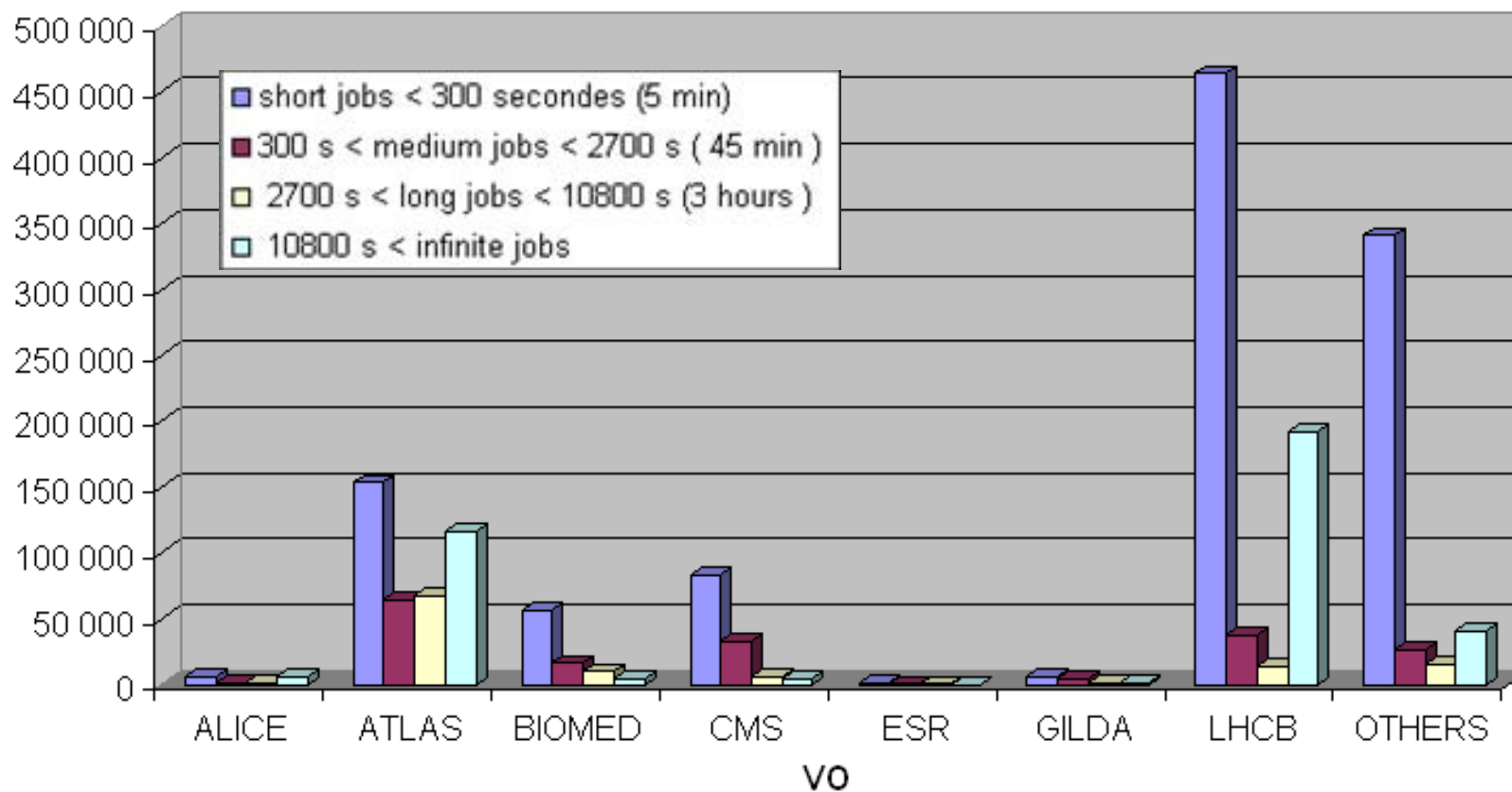
Elsewhere: 208 institutes, 1632 users

- Job submission per VO

<http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/durationbyVO.php>

Number of jobs

Done jobs duration distribution



- 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)

- New applications to join in the near future
 - Especially in the field of drug discovery


Enabling Grids for E-science

Welcome to the GENIUS INFN GRID Portal - Mozilla


File Edit View Go Bookmarks Tools Window Help


https://grid-tutor.ct.infn.it/


Home Bookmarks Red Hat, Inc. Red Hat Network Support Shop Products Training



INFN
Istituto Nazionale
di Fisica Nucleare







Enabling Grids for
E-science in Europe

Grid Enabled web eNvironment for site Independent User job Submission

RB: gilda	VO: gilda	RLS: GILDA	Your Data	Logout
-----------	-----------	------------	-----------	--------

Directory contents - tmp1100001761583.ef/gate_job_list_20041109_123955

RelDoseTree.gif (GIF Image, 606x302 pixels) - Mozilla

https://grid-tutor.ct.infn.it/ef/download/RelDose

Jobs Services

- up
- Jobs Settings
- Jobs Submission
- Jobs Queue
- GATE job data

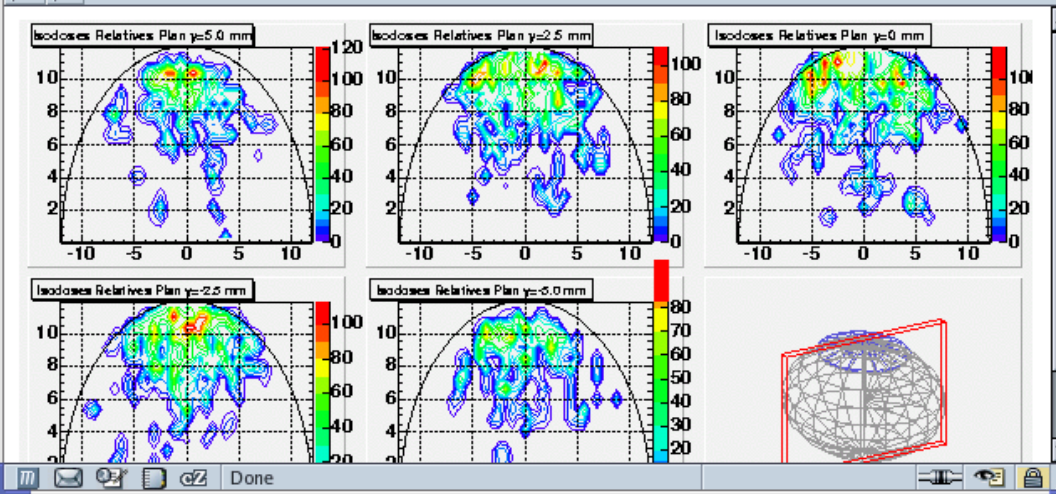
powered by
EnginFrame 3.2
compliant with
LCG-2
GRID.IT

ResultTOT

giorgio lo

RelDoseT

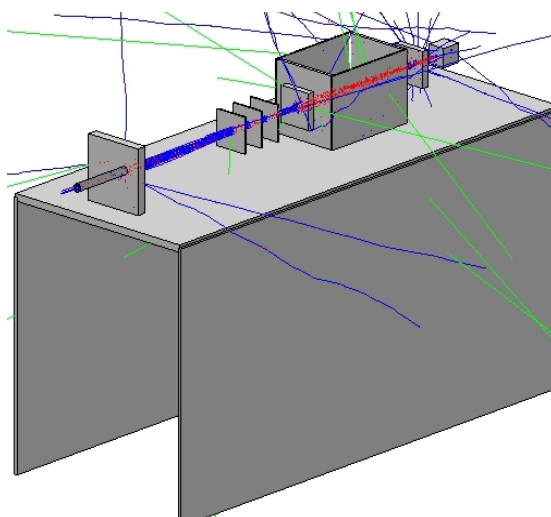
RelDoseT



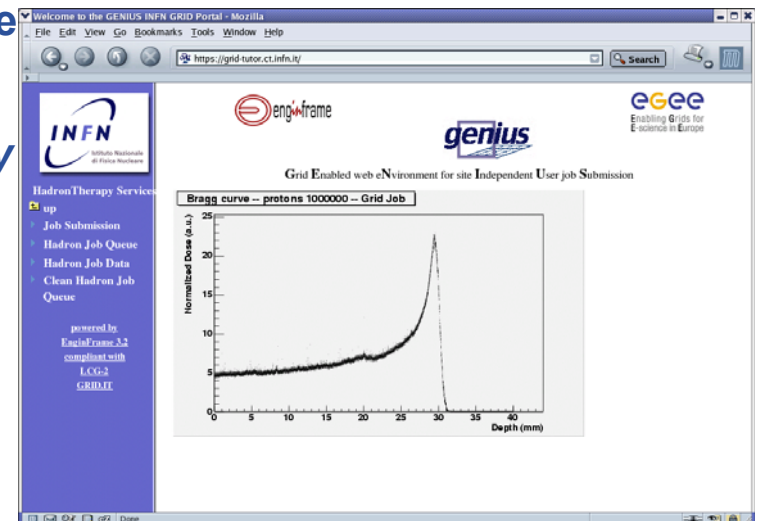
CATANA beam line in reality



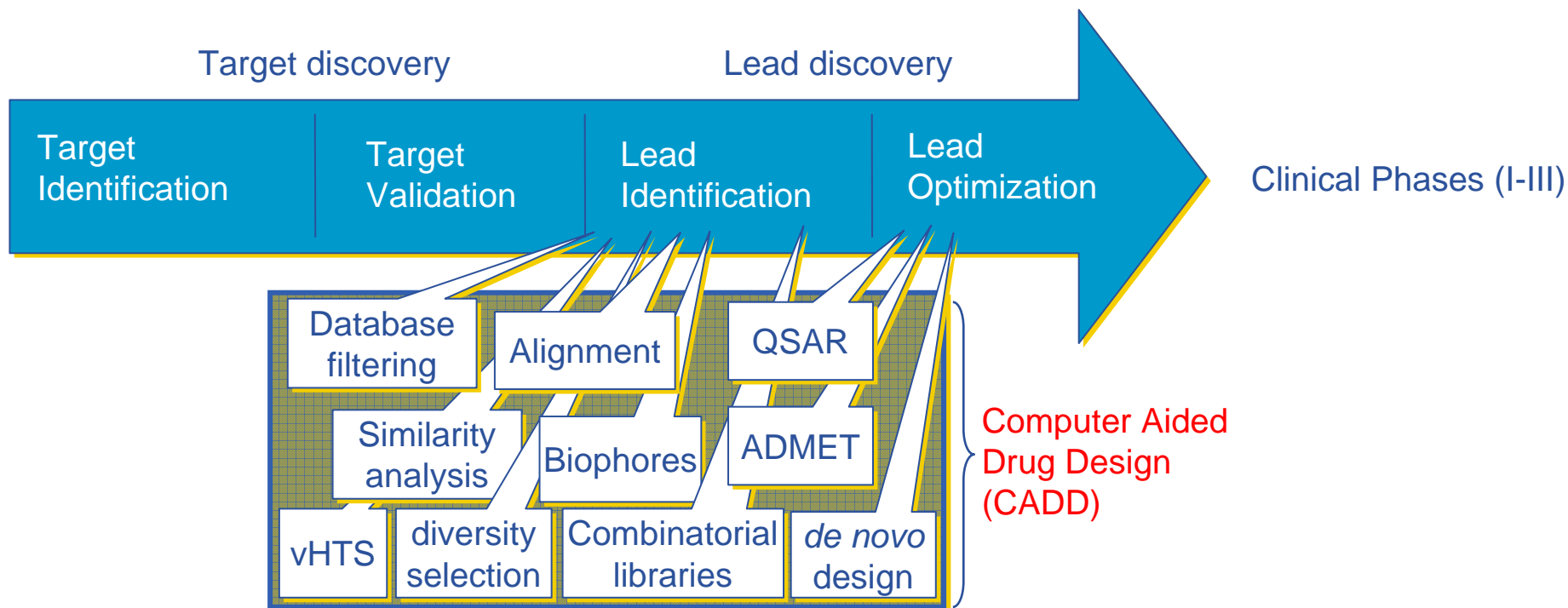
hadronTherapy in GENIUS



CATANA beam line simulated by hadronTherapy

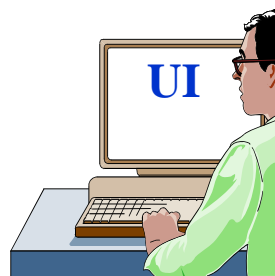


- **Demonstrate the relevance and the impact of the grid approach to address Drug Discovery for neglected diseases.**

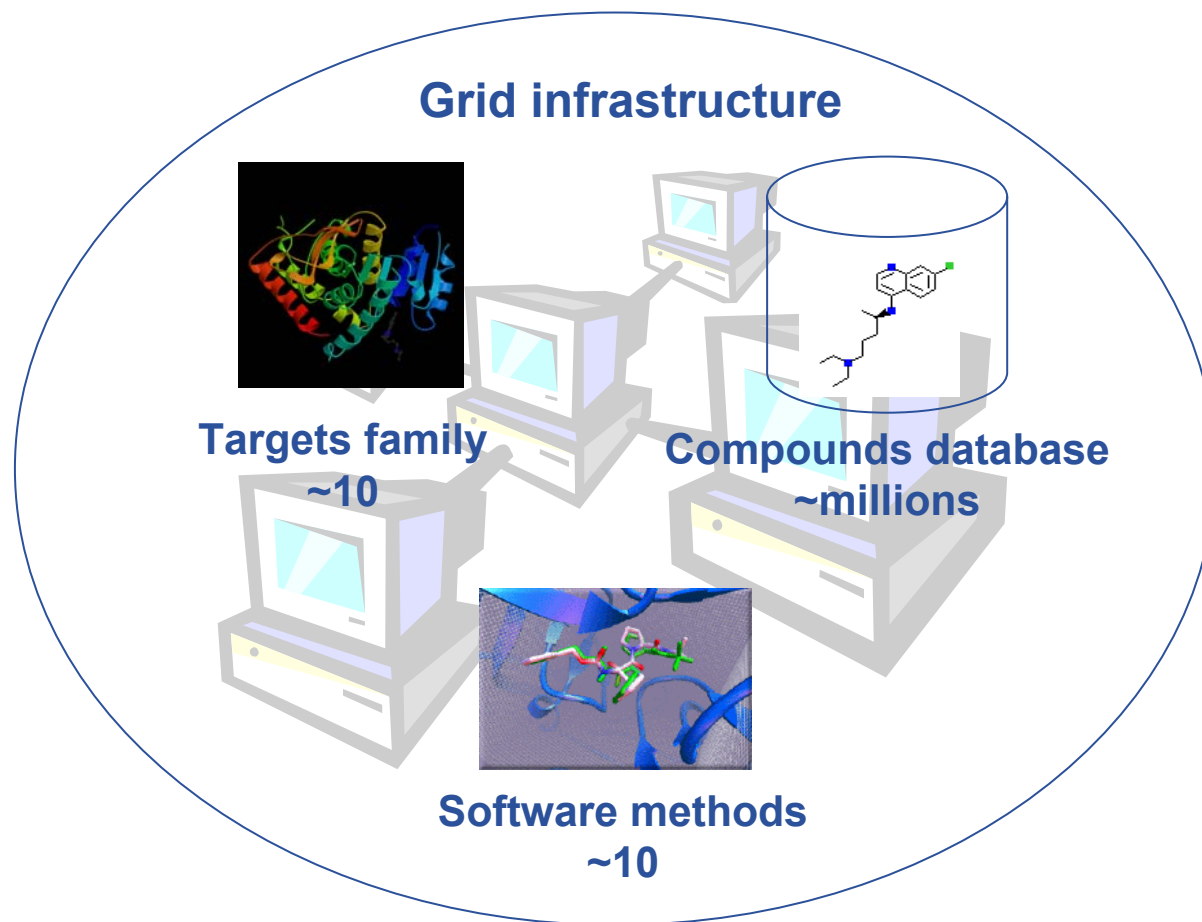


Duration: 12 – 15 years, Costs: 500 - 800 million US \$

- Predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure



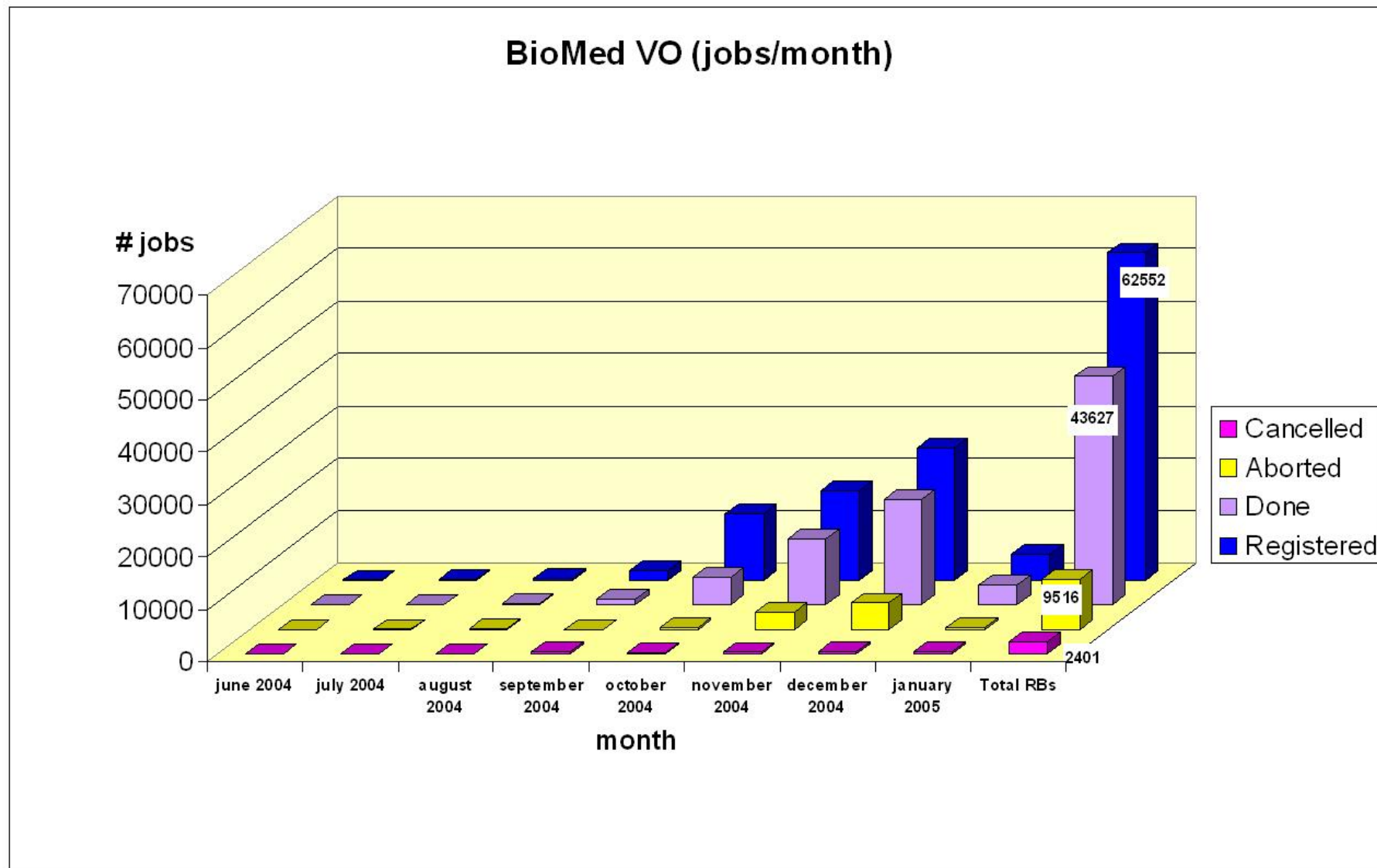
Parameter /
scoring settings



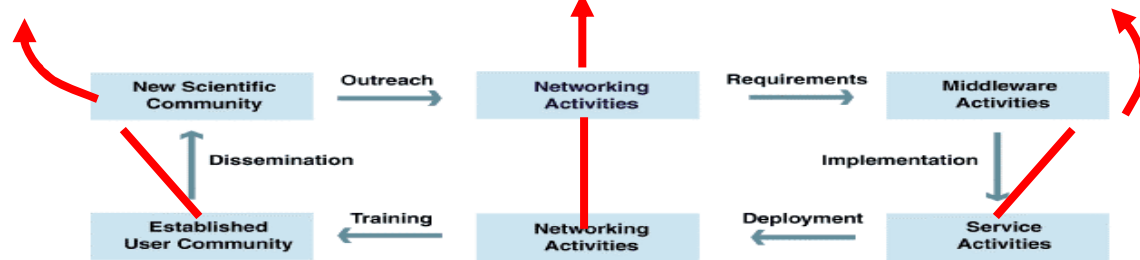
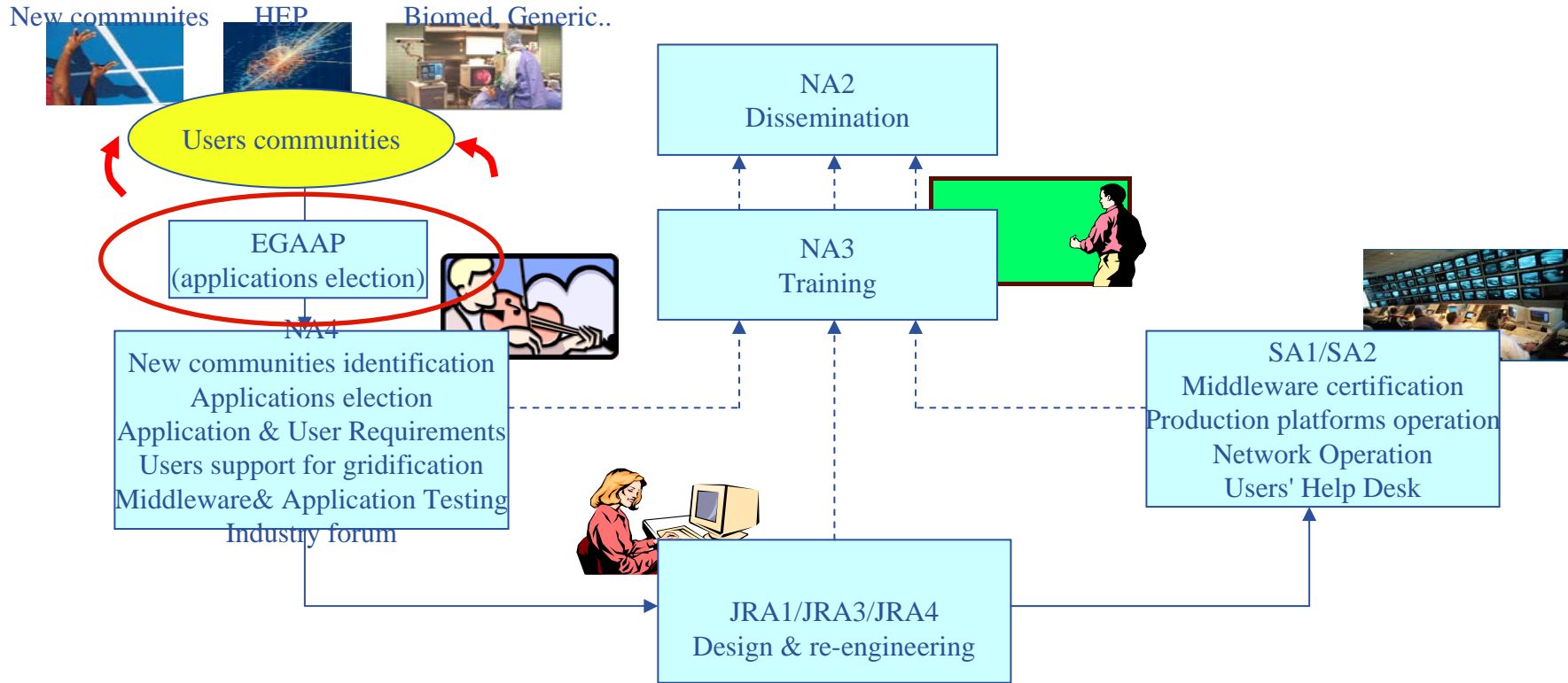
- **Malaria target sent by the inputSandbox**
 - Lactate dehydrogenase (Energy production, inhibited by chloroquine)
 - Default parameter / scoring settings

- **Compounds databases deployed on each SE**
 - NCI, National Cancer Institute compounds database
 - 2000 compounds
 - Ambinter, subset of ZINC : a free database of commercially-available compounds for virtual screening
 - 416 000 compounds, 3GB

- **Docking software**
 - Autodock : automated docking of flexible ligands to macromolecules
 - ~2,5 mn by target – compound job
 - Sent on each CE
 - FlexX : commercial prediction of protein-compound interactions
 - ~1mn by target – compound job
 - Available on SCAI node, soon on LPC node



The EGEE "Virtuous Cycle"



Activities mainly involved in the virtuous cycle

- **EGEE Generic Applications Advisory Panel** is the entry door for all 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

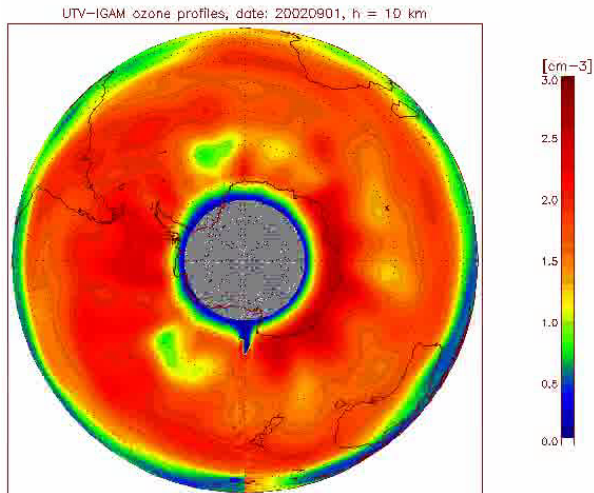
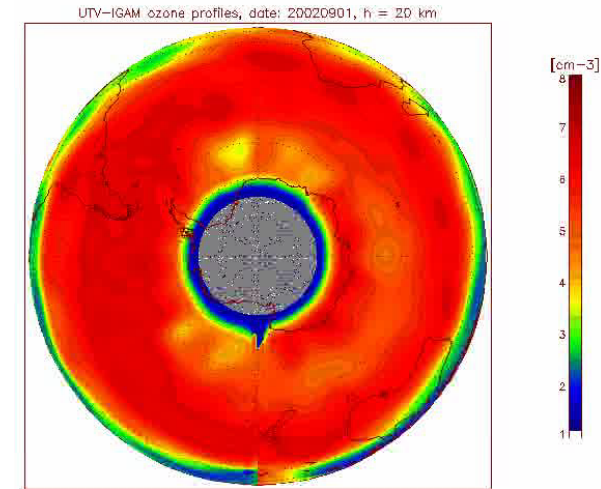
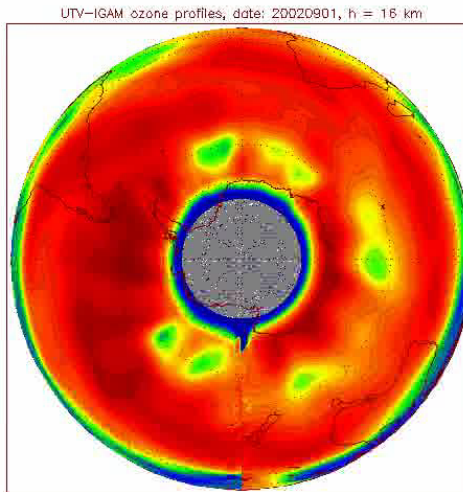
- **First call for proposals: limited distribution on May 17 2005**
- **First EGAAP meeting June 14 2004, at CERN**
 - 5 applications, 3 recommended for approval
 - Computational Chemistry
 - MAGIC, Astrophysics
 - Earth Science
- **EGAAP Recommendations approved by EGEE management on July 14.**
- **Second call for proposals : widest distribution possible on Sep 28, 2004**
- **Second EGAAP meeting November 25, Den Haag**
 - 7 applications received, 4 recommended for approval
 - Earth sciences (Solid Earth Physics, Hydrology)
 - Cosmology (Planck collaboration)
 - Drug discovery (Molecular docking)
 - Search engines for digital knowledge (GRACE)
- **Recommendations approved by EGEE management on 17 December 2004.**
- **Third EGAAP meeting April 21, Athens**
 - 3 applications received, 2 recommended for approval
 - Drug Discovery
 - E-Grid (e-finance, e-business)

A photograph of the ENVISAT satellite in orbit above Earth. The satellite is a complex, gold-colored structure with various instruments and antennas. A long, blue solar panel array extends from the satellite. The Earth's surface is visible below, showing blue oceans and white clouds.

ENVISAT

- 3500 Meuro programme cost
- Launched on February 28, 2002
- 10 instruments on board
- 200 Mbps data rate to ground
- 400 Tbytes data archived/year
- ~100 'standard' products
- 10+ dedicated facilities in Europe
- ~700 approved science user projects

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



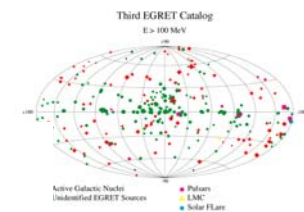
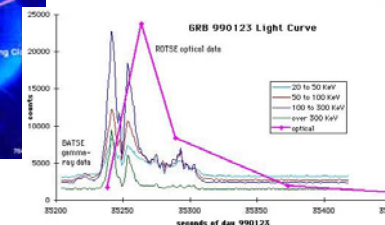
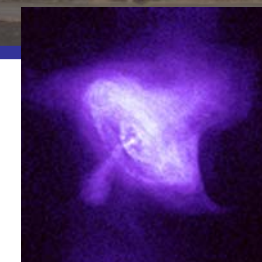
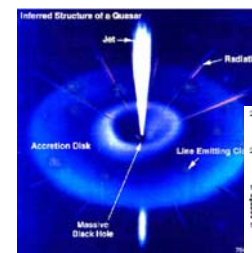
- Ground based Air Cerenkov Telescope
- Gamma ray: 30 GeV - TeV
- LaPalma, Canary Islands (28° North, 18° West)
- 17 m diameter
- operation since autumn 2003 (still in commissioning)
- Collaborators:

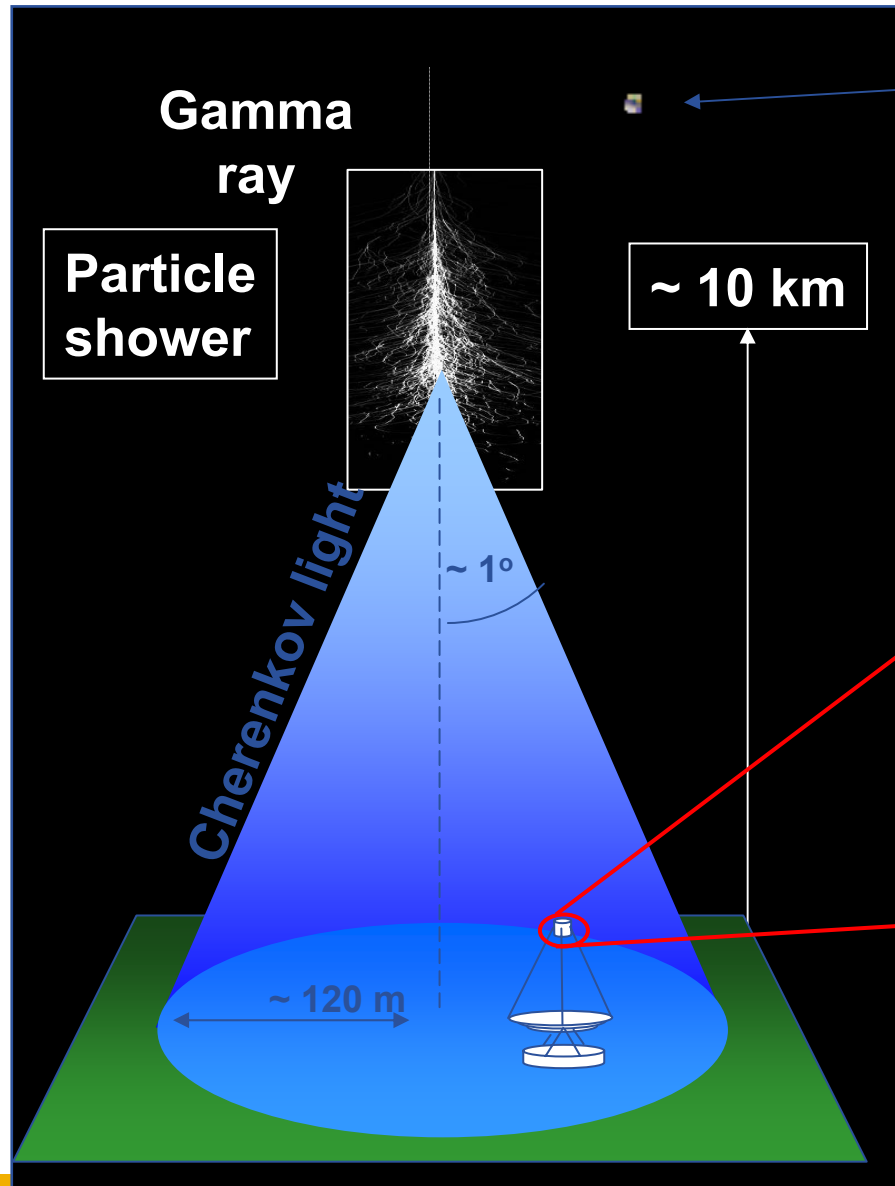
IFAE Barcelona, UAB Barcelona, Humboldt U. Berlin, UC Davis, U. Lodz, UC Madrid, MPI München, INFN / U. Padova, U. Potchefstroom, INFN / U. Siena, Tuorla Observatory, INFN / U. Udine, U. Würzburg, Yerevan Physics Inst., ETH Zürich



Physics Goals: Origin of VHE Gamma rays

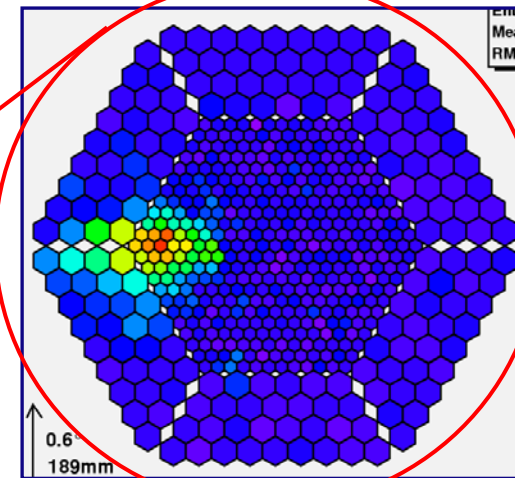
- Active Galactic Nuclei
- Supernova Remnants
- Unidentified EGRET sources
- Gamma Ray Burst





GLAST ($\sim 1\text{ m}^2$)

Cherenkov light Image of particle shower in telescope camera



reconstruct:
 arrival direction, energy
 reject hadron background

- **Based on extensive Monte Carlo Simulation**
 - air shower simulation program CORSIKA
 - Simulation of hadronic background is very CPU consuming
 - to simulate the background of one night, 70 CPUs (P4 2GHz) needs to run 19200 days
 - to simulate the gamma events of one night for a Crab like source takes 288 days.
 - At higher energies (> 70 GeV) observations are possible already by On-Off method (This reduces the On-time by a factor of two)
 - Lowering the threshold of the MAGIC telescope requires new methods based on Monte Carlo Simulations

- **History**

- The

- **Now:**

- Resources

```
bash-2.05a$ lcg-infosites --vo magic all
LRC endpoint for magic: http://rlsmagic.pic.es:8080/magic/edg-local-replica-catalog/services/edg-local-replica-catalog
RMC endpoint for magic: http://rlsmagic.pic.es:8080/magic/edg-replica-metadata-catalog/services/edg-replica-metadata-catalog
LRC endpoint for magic: http://rlsmagic.pic.es:8080/magic/edg-local-replica-catalog/services/edg-local-replica-catalog
RMC endpoint for magic: http://rlsmagic.pic.es:8080/magic/edg-replica-metadata-catalog/services/edg-replica-metadata-catalog
```

These are the related data for magic: (in terms of CPUs)

#CPU	Free	Total	Jobs	Running	Waiting	Computing	Element
20	20	0	0	0	0	0	ce01.pic.es:2119/jobmanager-torque-magic
168	4	1	1	0	0	0	lcgce02.ifaes:2119/jobmanager-lcgpbs-magic
7	6	0	0	0	0	0	ce001.grid.bas.bg:2119/jobmanager-lcgpbs-magic
82	63	6	6	0	0	0	gridba2.ba.infn.it:2119/jobmanager-lcgpbs-long
82	63	3	3	0	0	0	gridba2.ba.infn.it:2119/jobmanager-lcgpbs-short
1280	5	0	0	0	0	0	lcg-gridka-ce.fzk.de:2119/jobmanager-pbspro-long
1280	5	0	0	0	0	0	lcg-gridka-ce.fzk.de:2119/jobmanager-pbspro-short
50	29	1	1	0	0	0	grid012.ct.infn.it:2119/jobmanager-lcgpbs-long
50	29	0	0	0	0	0	grid012.ct.infn.it:2119/jobmanager-lcgpbs-short
82	63	8	8	0	0	0	gridba2.ba.infn.it:2119/jobmanager-lcgpbs-infinite
1280	5	0	0	0	0	0	lcg-gridka-ce.fzk.de:2119/jobmanager-pbspro-default
582	413	0	0	0	0	0	ce01-lcg.cr.cnaf.infn.it:2119/jobmanager-lcglsf-magic
1280	5	0	0	0	0	0	lcg-gridka-ce.fzk.de:2119/jobmanager-pbspro-extralong
50	29	24	20	4	0	0	grid012.ct.infn.it:2119/jobmanager-lcgpbs-infinite
8	6	1	1	0	0	0	gridit-ce-001.cnaf.infn.it:2119/jobmanager-lcgpbs-long
8	6	0	0	0	0	0	gridit-ce-001.cnaf.infn.it:2119/jobmanager-lcgpbs-short
8	6	1	1	0	0	0	gridit-ce-001.cnaf.infn.it:2119/jobmanager-lcgpbs-infinite

The total values are:

```
-----
4889 648 45 41 4
```

These are the related data for magic: (in terms of SE)

Access point	Avail Space(Kb)	Used Space(Kb)	Files	SEs
gsiftp://castorgrid.ifaes/castor/ifaes/lcg/magic/	100000000000	0	0	castorgrid.ifaes
gsiftp://castorgrid.pic.es/castor/ifaes/lcg/magic/	100000000000	0	0	castorgrid.pic.es
gsiftp://castorgrid.ifaes/castor/ifaes/lcg/magic/	53020932	0	0	castorgrid.ifaes
gsiftp://se001.grid.bas.bg/storage/magic/	354813480	30087	4	se001.grid.bas.bg
gsiftp://gridba6.ba.infn.it/flatfiles/SE00/magic/	1605498664	0	0	gridba6.ba.infn.it
gsiftp://grid007g.cnaf.infn.it/flatfiles/SE00/magic/	1497769448	0	0	grid007g.cnaf.infn.it
gsiftp://lcg-gridka-se.fzk.de/grid/fzk.de/.../SE00/magic/	1605498664	0	0	lcg-gridka-se.fzk.de



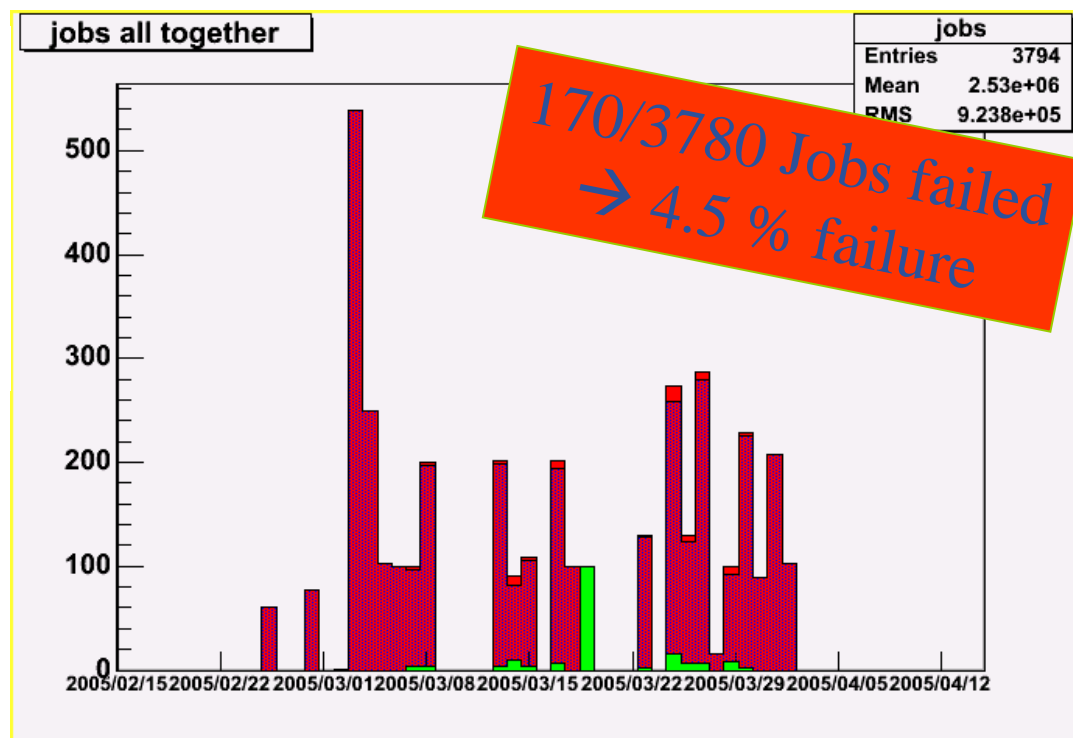
Data challenge Grid-1

- 12M hadron events
 - 12000 jobs needed
- started march 2005
- up to now ~ 4000 jobs

Job successful:
Output file
registered at PIC

First tests:

- with manual GUI submission
- Reasons for failure
 - Network problems
 - RB problems
 - Queue problems

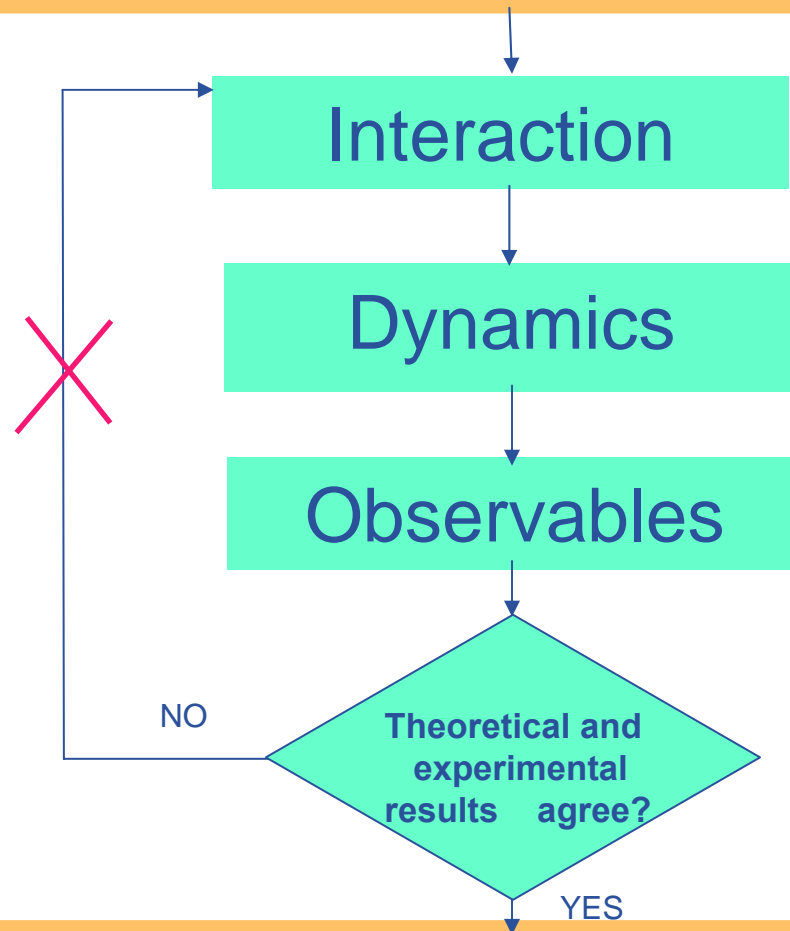


Diagnostic:

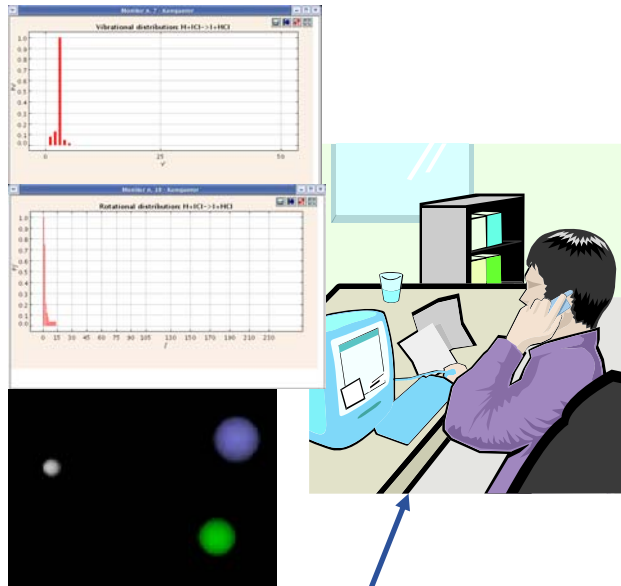
- no tools found
- complex and time consuming
- use metadata base, log the failure,
resubmit and don't care

Organization	Actor	Contact email
University of Barcelona	Margarita Alberti	maw@qf.ub.es
University of the Basque Country	Ernesto Garcia	qfpgapae@vc.ehu.es
Hungarian Academy of Sciences	Gyorgy Lendvay	lendvay@chemres.hu
University of Geneva	Laura Gagliardi	laura.gagliardi@unipa.it
IMIP- Italian National Research Council	Mario Capitelli	m.capitelli@area.ba.cnr.it
University of Goteborg	Gunnar Nyman	nyman@phc.gu.se
University of Crete	Stavros C Farantos	farantos@iesl.forth.gr
Swiss Federal Institute of Technology (ETH)	Hans Peter Luethi	hans.luethi@sl.ethz.ch

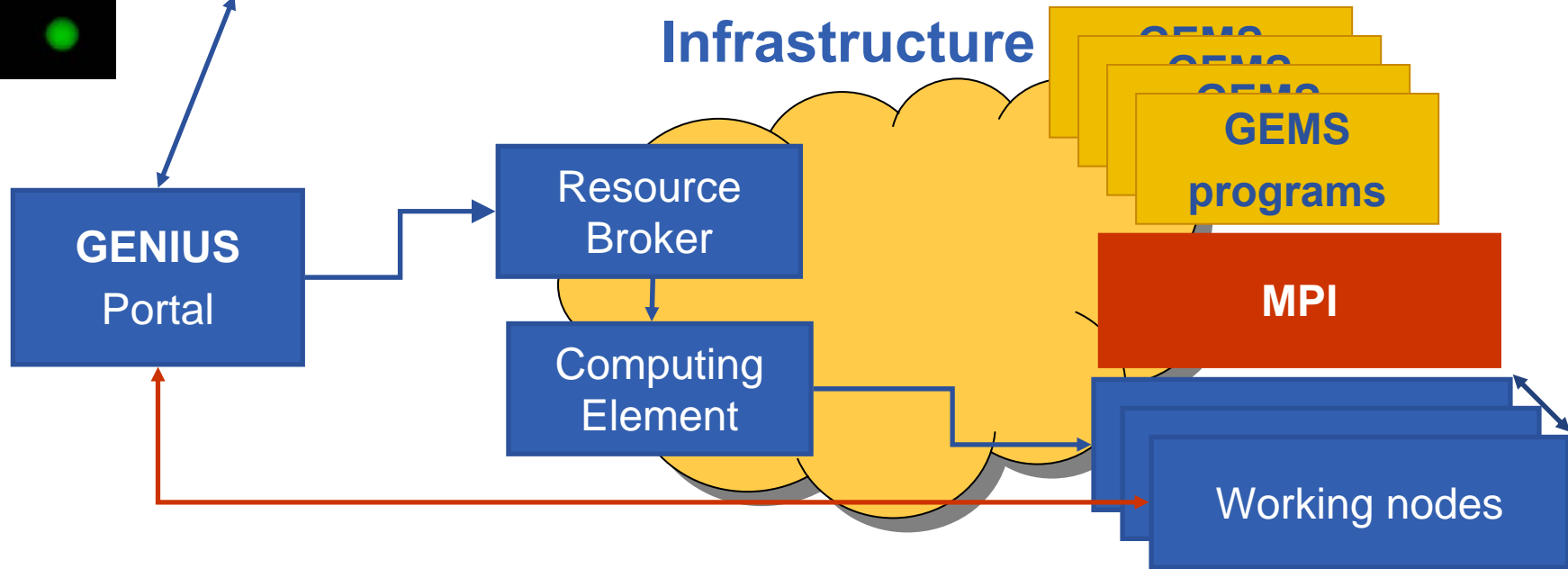
REQUEST: a potential fitted to beam experiments



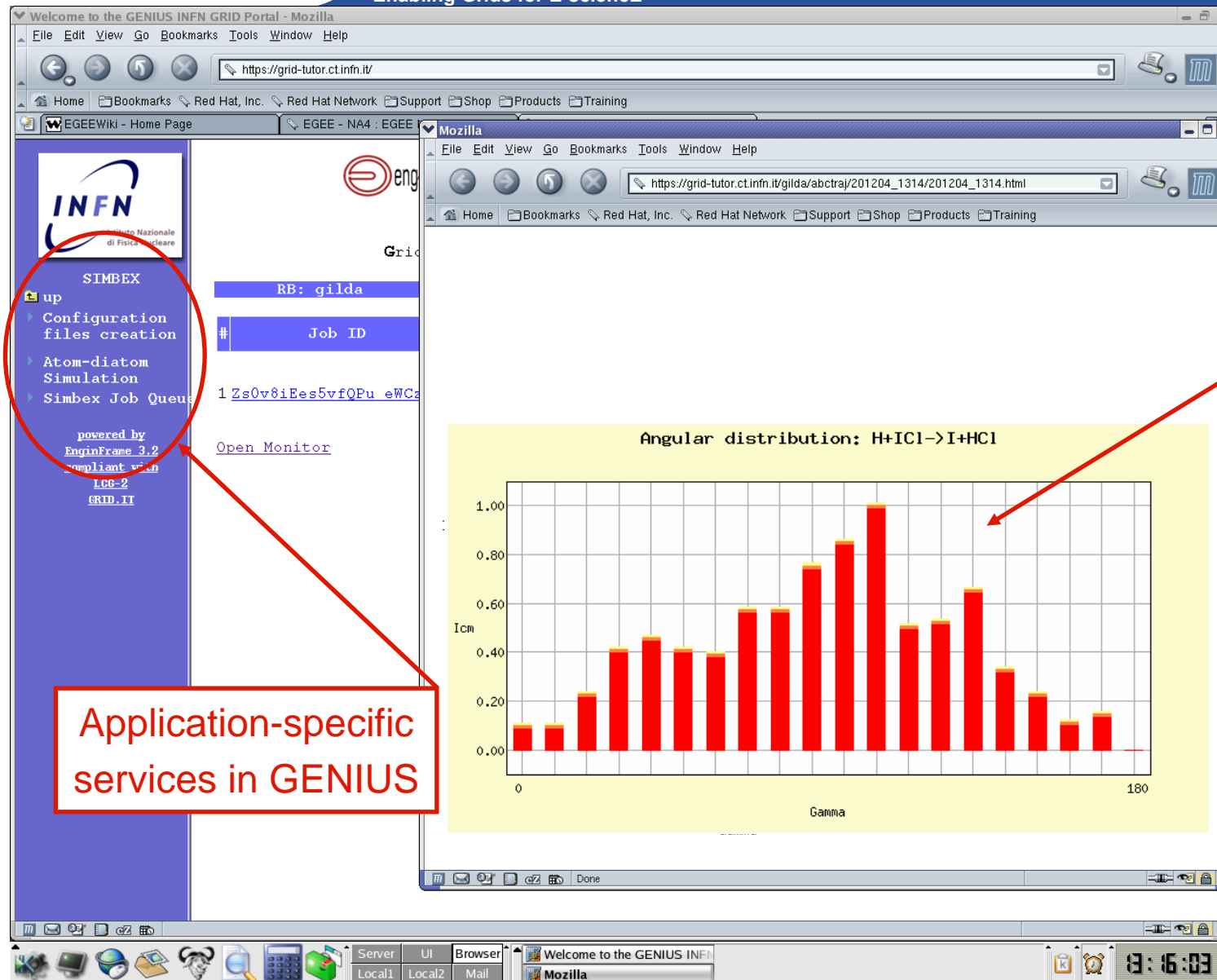
SUPPLY: the potential and related monitors



GILDA Testbed Infrastructure



In-Out-bound connectivity



powered by EnginFrame 3.2

compliant with LCG-2 GRID.IT

RB: gilda

#	Job ID
1	Zs0v8iEes5vfQPu_eWCz

Open Monitor

Angular distribution: H+ICI → I+HCI

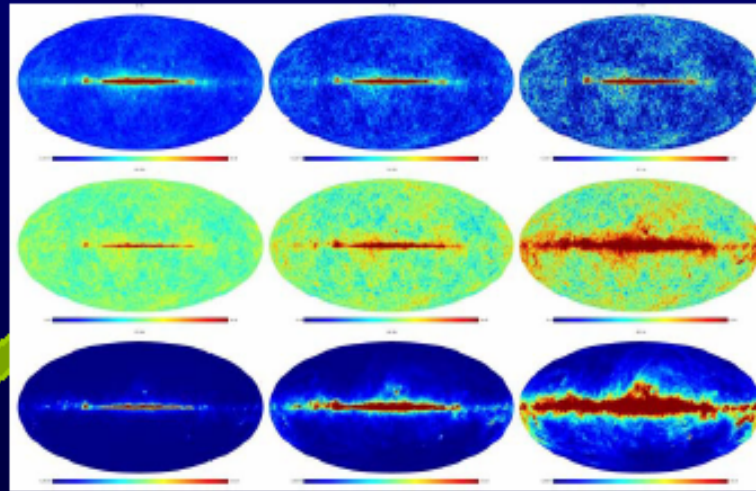
Gamma

Interactive
MPI jobs !

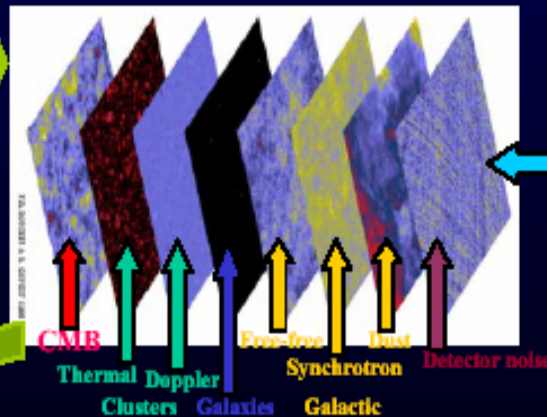
Application-specific
services in GENIUS

Date	Milestone	Task
		Gilda Deployment
		Application development
Sept 30, 2005	GEMS.1 deployment on the production grid	Deployment on EGEE
Nov 30, 2005	GEMS.1 tested	Test
Jan 31, 2006	GEMS.1 on production	Production
Jan 31, 2006	GEMS.2 development	Application development
Mar 30, 2006	Metrics of satisf. of GEMS.1	Metrics of satisfaction
Apr 30, 2006	GEMS.2 deployment on the production grid	Deployment on EGEE
Jul 31, 2006	GEMS.2 tested	Test
Oct 31, 2006	Metrics of satisf. of GEMS.2	Metrics of satisfaction

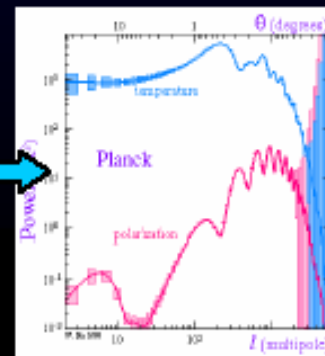
PLANCK



*Imaging the sky emission
at many frequencies*

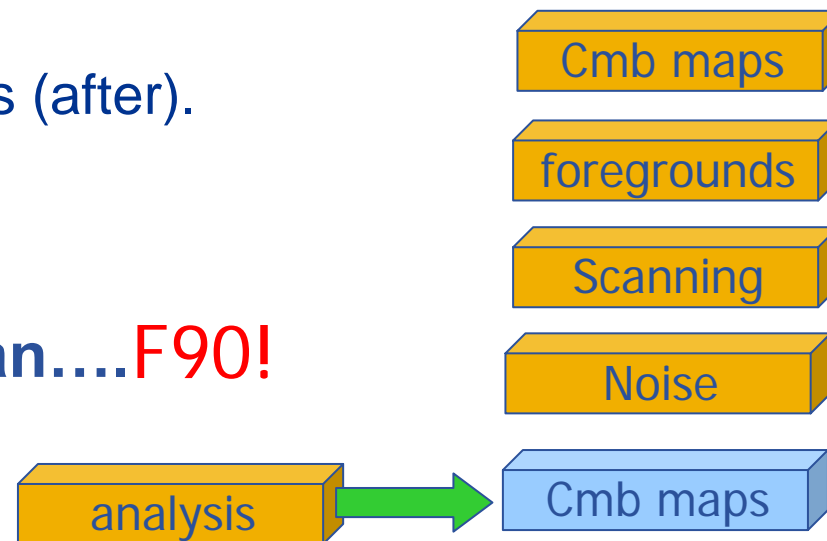


Peeling back the layers



*Recovering the
cosmological information*

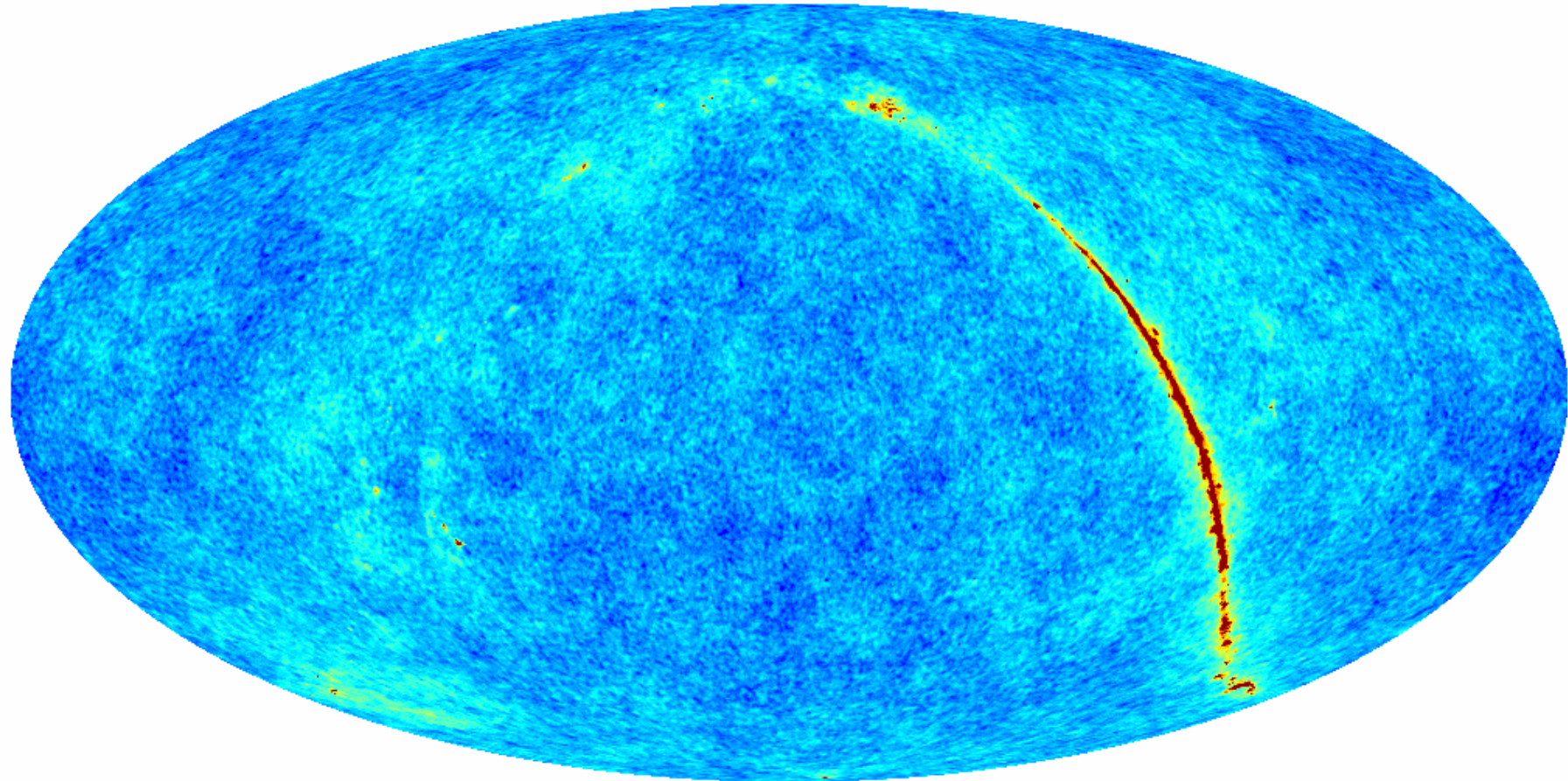
- **Mission simulation Software:**
 - ground checks (before);
 - control check & corrections (after).
- **Pipeline:**
 - Chained but not parallel;
- **Stages are C/C++/Fortran....F90!**
- **Shell/perl scripts;**



Planck simulation is a set of 70 instances of the Pipeline.

- **Is it parallel? NO it runs concurrently.**
- **Do we need MPI/parallel? Yes. In later phase for data analysis (16/32 CPUs in the site).**
- **How long does it run? From 6h up to 36h**
- **Do we produce data? YES, we have an intensive data production. Can be more than 1 PB.**
- **Access/exchange data from other experiments (IVOA, MAGIC)**

Synthesized Sky Map LFI 70 GHz



1.381100



1.382900

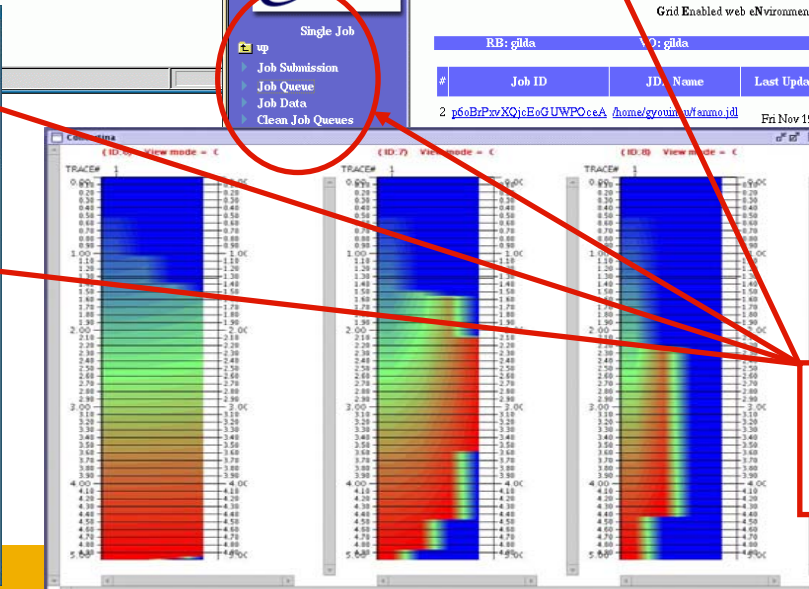
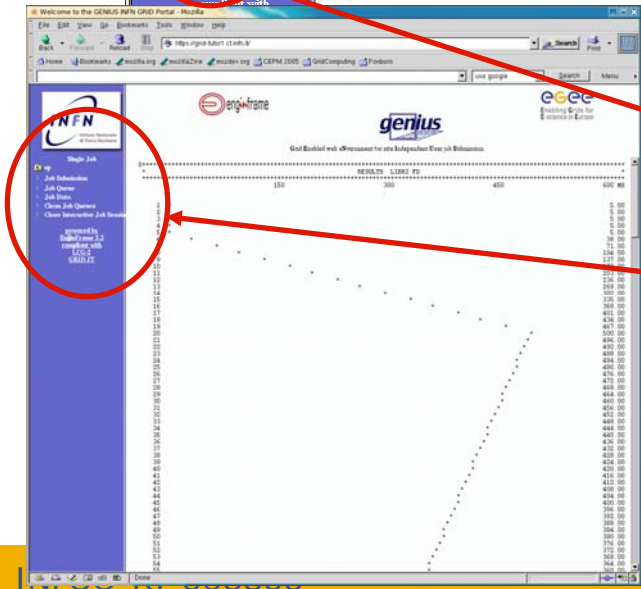
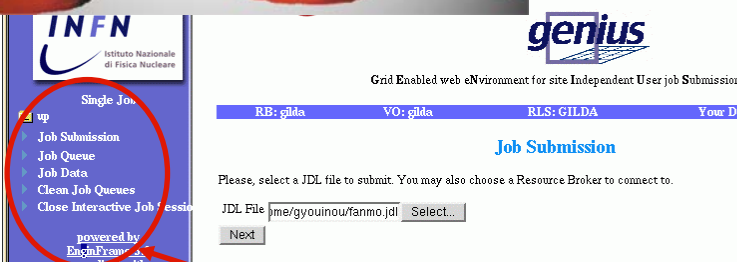
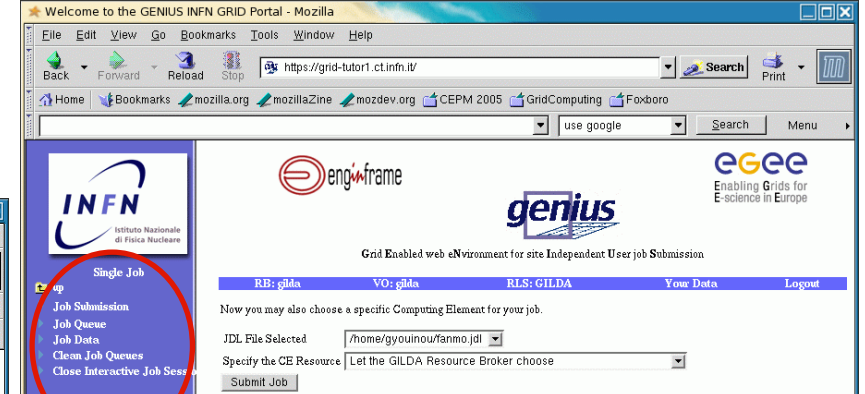
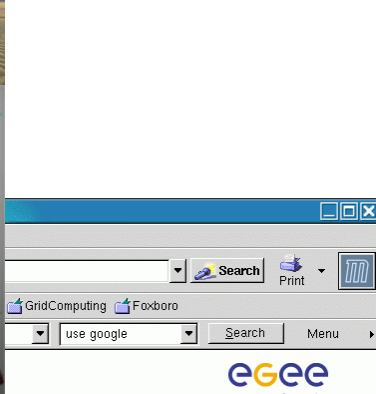
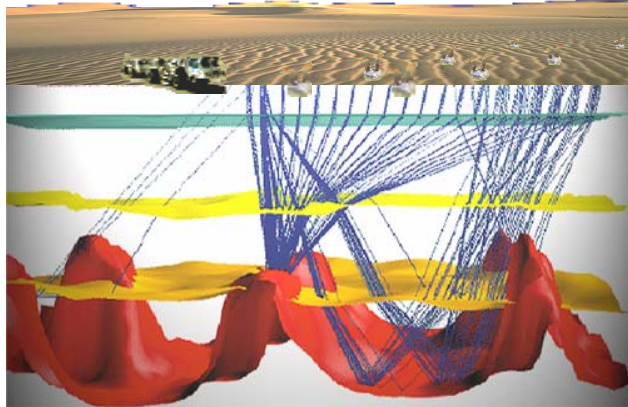
- **Objectives:**
 - To promote and disseminate Grid concepts towards industry and service groups
 - To raise the awareness of EGEE within industry
 - To encourage businesses to participate in the project
- **Members:** interested companies having activities in Europe
- **Management:**
 - Steering Committee
 - Representatives of EGEE partners
 - Representatives of main industrial sectors

MICHELIN	ACRI-ST	
CEA	TOTAL	
GRIDSYSTEMS	IFP	Pôle Européen Plasturgie
CREDIT LYONNAIS	C-S	SCAI
CNES	DATA SYNAPSE	GENIAS Benelux
BNP	THALES	Arcelor
PALLAS-INTEL	LION Bioscience AG	Fugitsu
DASSAULT AVIATION	BULL	IBM
EADS CCR	Sanofi-Aventis	HUTCHINSON
SNECMA	SOCIETE GENERALE	Gridwise Technologies
STMicroelectronics Srl	NICE	Compagnie Générale de Géophysique
Hewlett-Packard	NOVARTIS PHARMA AG	SCHLUMBERGER
CSTB	EDF	Datamat
Daimler	ESI	AGENIUM Technologies
Oracle	Sun	HLRS
PSA	ICATIS	Telefonica Spain
PLATFORM COMPUTING	CERFACS	AIRBUS
GRIDXPRT	MICROSOFT	DUTCH SPACE paris Office
PECHINEY CRV	British Telecom	ORION LOGIC Ltd.
France Télécom	CENAERO	CSCS



Enabling Grids for E-science

Earth Science (industrial): EGEODE example

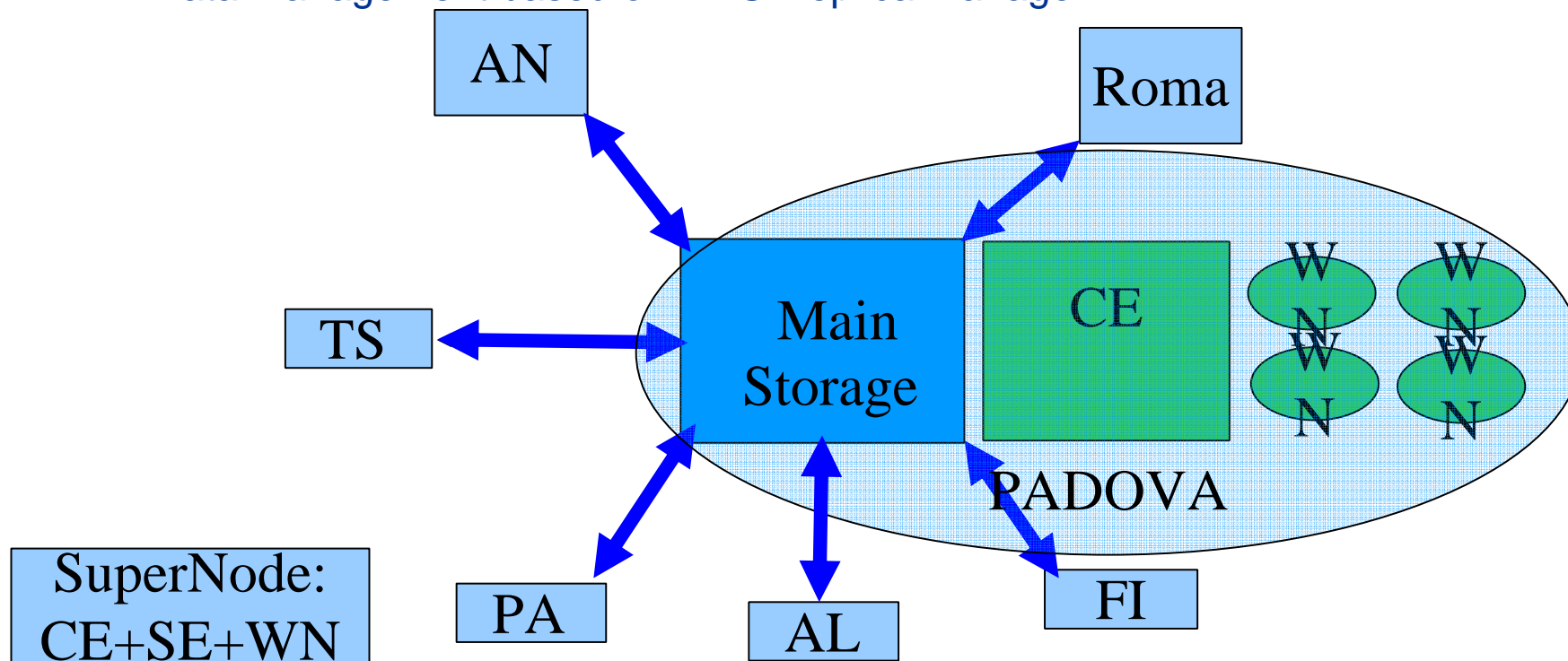


Application-specific services in GENIUS

- **A FISR project funded by MIUR (2003)**
- **Original Aim:**
 - Use of grid paradigm for research on complex system in economics and finance
- **MIUR asked to similar FISR projects to coordinate together**
- **ICTP new task:**
 - Provide a National Italian Facility for economical and financial data using grid technology

- **Several different research groups in Italy**
 - Palermo: dep of Physics;
 - Roma: dep. Of Physics;
 - Roma: IAC-CNR Istitute;
 - Ancona: Department of Economics;
 - Firenze: dep. of Statistics.
 - Alessandria: dep of Physics;
 - Trieste:
 - *Dep of Economics*
 - *Dep of Engineering*
 - *Area Science Park*
 - *ICTP*
- **2 groups in USA: UCLA + Columbia**
 - Grid research applied to finance

- Version 1.0 released: 8-9 october 2004 (meeting in Trieste)
- Globus based EDG middleware.
- Specific VO within Grid.it. (our Grid provider..)
- A star shaped topology for data management:
 - Central SE (large and backed up storage) hosted in grid.it
 - Local SE hosted in research centres acting as a temporary storage space.
 - Data management based on EDG Replica manager



- **EGEE is the first attempt to build a worldwide Grid infrastructure for data intensive applications from many scientific domains**
- **A large-scale production grid service is already deployed and being used for HEP and BioMed applications with new applications being ported**
- **Resources & user groups will rapidly expand during the project**
- **A process is in place for migrating new applications to the EGEE infrastructure**

