Introduction to Grids & e-Science

Dr. Isabel Campos Plasencia

Científico Titular

Instituto de Física de Cantabria

General Introduction

- **Grid computing**
  - **Definition, technological context, evolution,…**
- **Grids for scientific computing**
  - Elements of a Grid Infrastructure
  - E-Science
- **Scientific applications on the Grid**

UIMP
Universidad Internacional Menéndez Pelayo

e-Science and e-Infrastructures

e-Science
e-Infrastructures
Grid
Aplications

IFCA
Instituto de Física de Cantabria

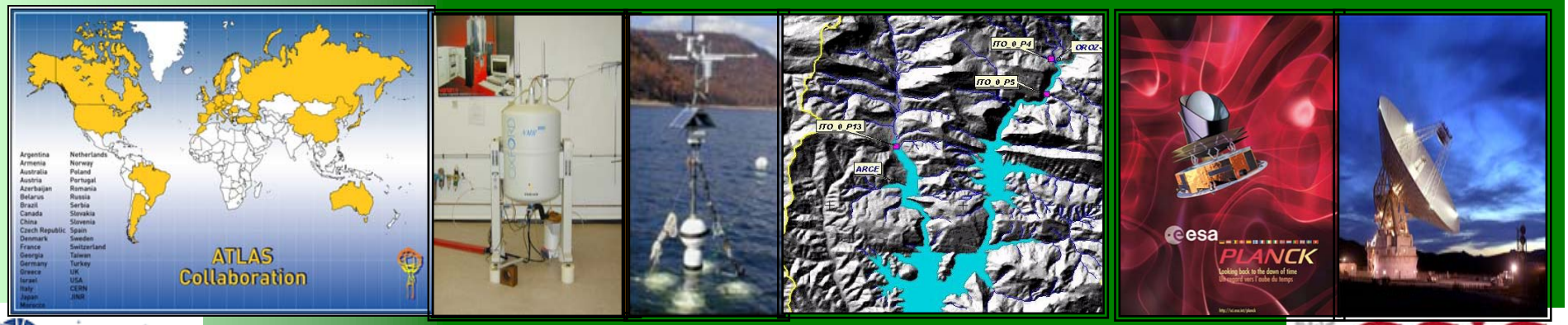
CSIC
Consejo Superior de Investigaciones Científicas

e-Science and e-Infrastructures: Definition

E-Infrastructure new generation of research infrastructures based on information and communication technologies

E-Science refers to scientific activities that are carried out by using resources distributed across the internet

- The utilization of those distributed resources is both a **necessity and an added value**
- More effective when associated to a global collaboration more than at the individual level



e-Science and e-Infrastructures: Definition

New concepts but

- The basics of scientific work is still the same
  - Observation, experiment,…
  - Analysis, Result Validation, Publication, discussion,…
    - In all the steps computing technologies are a key issue
    - As important in Physics as experimental observation

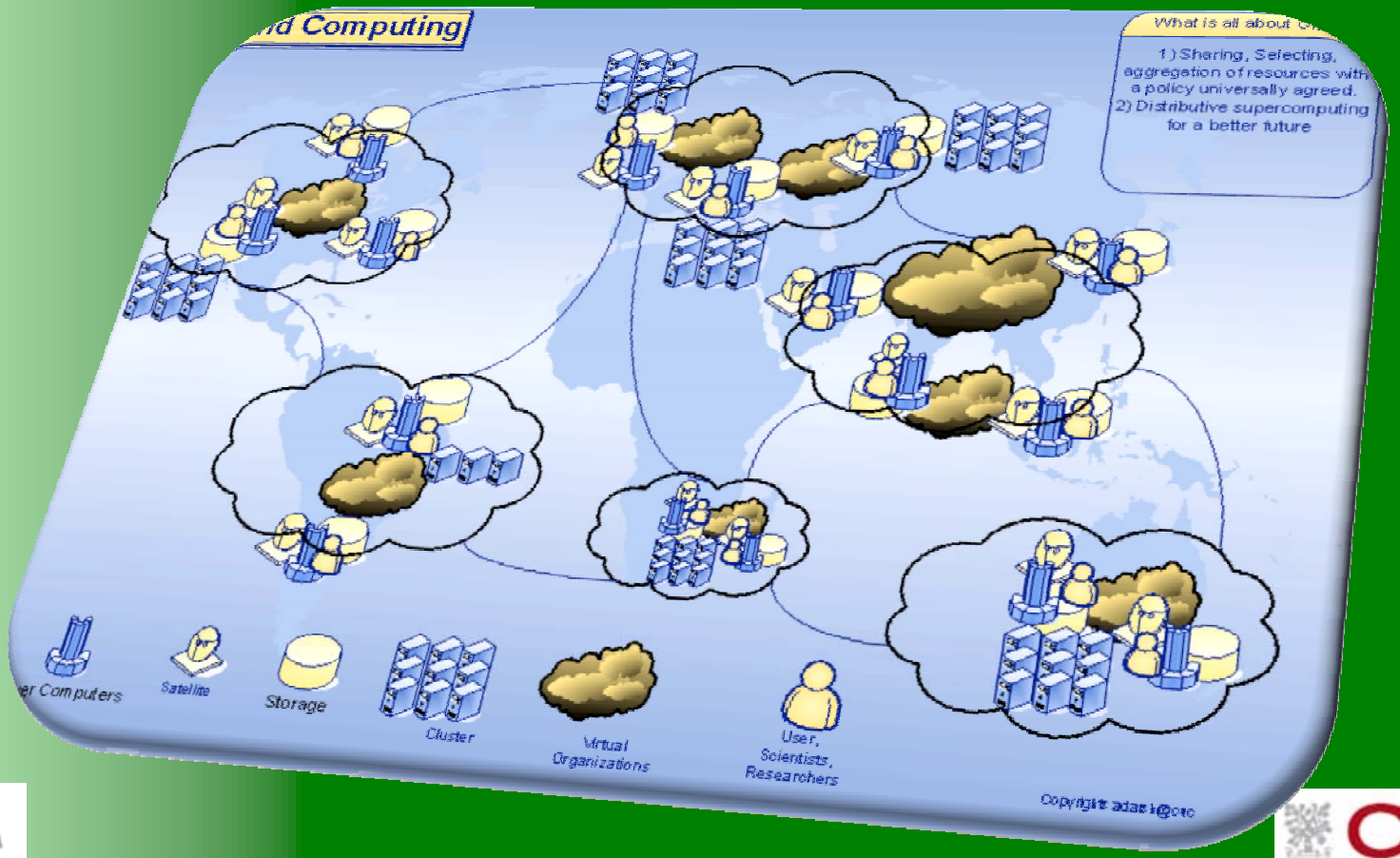
Observation, Experiment: Complex detectors located in accelerators, cameras installed in satellites, deployment of sensors networks for Earth Observation,…

Analysis, Modelling Result Validation: Computing resources

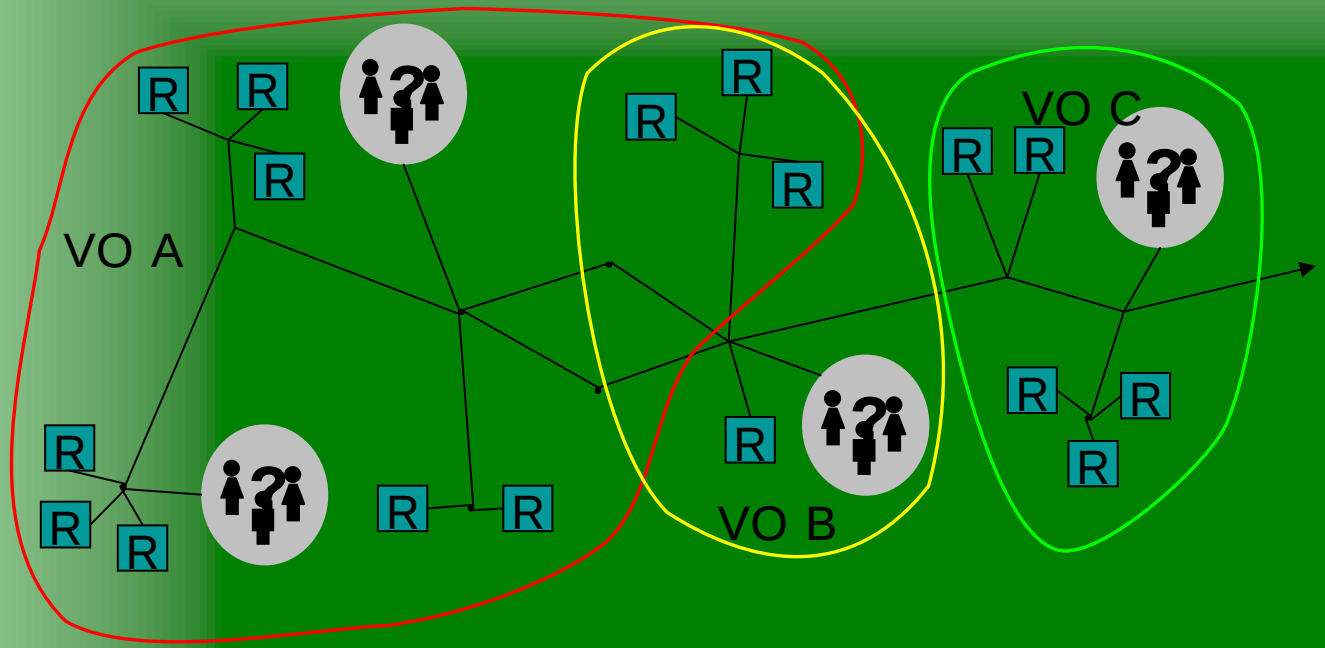
Advanced Computing Projects require furthermore: Specialized Hardware and Software; Methodology and Algorithmic developments

What is a Grid ?

A Grid is a set of resources, (digital instruments and elements attached to them or stored in them) which can be used in a combined way through a middleware to solve efficiently a particular problem

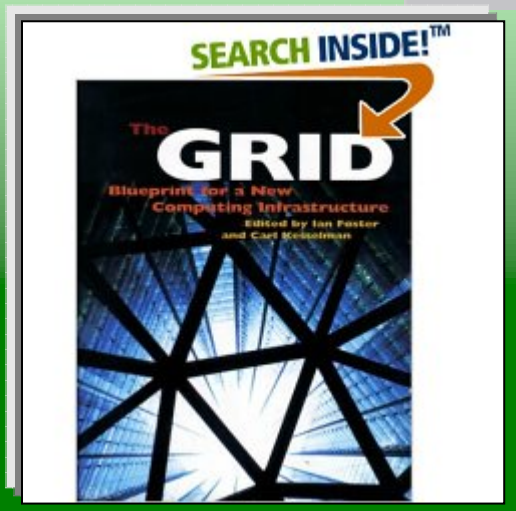


Grid and Virtual Organizations



Individual and organizations with a common interest Can benefit from sharing their computing resources Using Grid techniques. These groups of users are called Virtual Organizations.

How the Grid was born ?



- **"The Grid", Chapter 1,**
  (Ian Foster and Carl Kesselman, 1998)

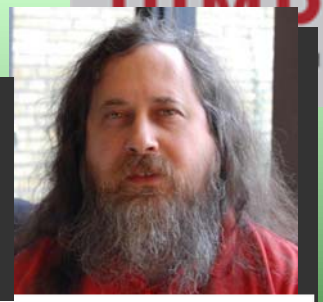
  The current state of computing would be analogous to the state of development of electricity power in the beginning of the XX century. The true revolution came not from the production of electricity but from thepossibility of distributing it transparently over a network.

- **Access to computational services (data storage and CPU) should be as transparent as using a plug**

  - The user does not need to know where his electricity has been generated
  - The user pays for consumption

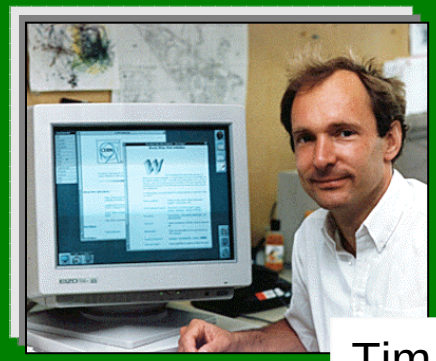
The technological context

- **Around 1998 we witnessed the boom of commodity computers**
  - Low priced hardware thanks to massive production of home PCs
  - Communication Networks become faster and more reliable;
  - Internet is starting to become the source of global information.
  - Linux operating system and in general GNU software are dominating the scientific computing world
    - **Linux clusters start to spread**
- **Scientist have the idea of developing a computing technology analogous to the web**
  - Sharing CPU instead of sharing information
  - **World Wide Web** has been developed middle of 90s at CERN
  - One of the most prominent examples of technology transfer from science to society



Richard Stallman

Linus Torvalds



Tim Berners Lee

The technological context: Globus

- **The Globus project** ***Globus Alliance: http://www.globus.org***
  - Development of software components to implement the philosophy of Grid computing (Open Source)
    - First release 1998 Globus 1.0
    - Latest release 2008 Globus 4.2.1
- **Globus is a set of software services and libraries to make possible**
  - Secure management of distributed infrastructures
  - Access to distributed resources and data
- **Developing an environment oriented to users and which allows to access remote resources in a controlled and secured way**

Elements of an e-Science Grid

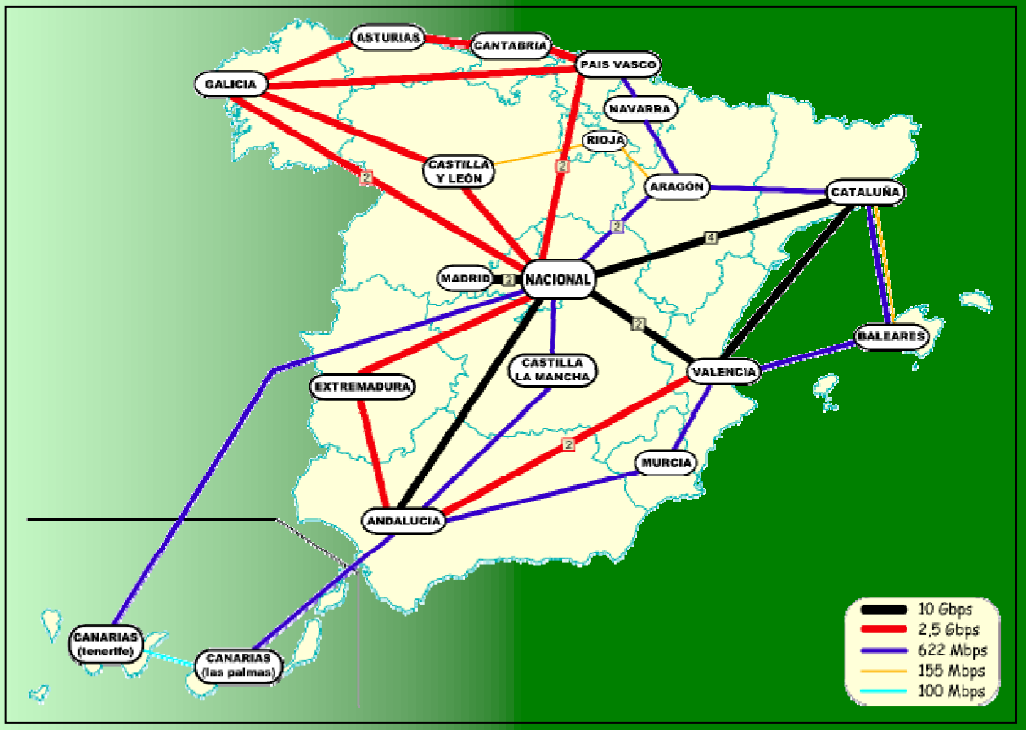
Network

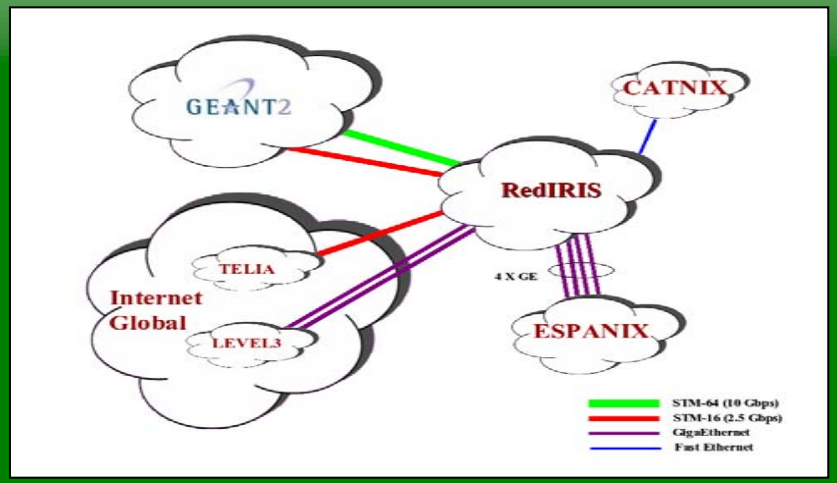
Hardware

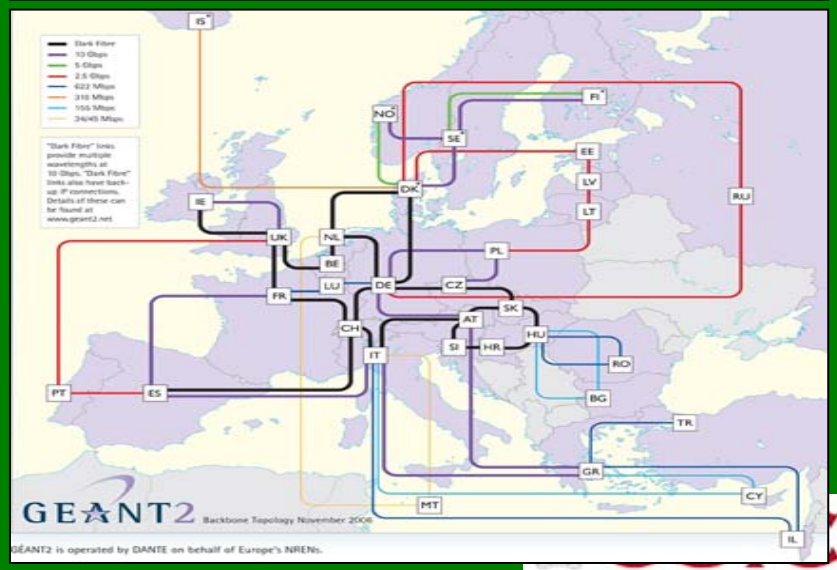
Software

Elements of a Grid

❑ The Network

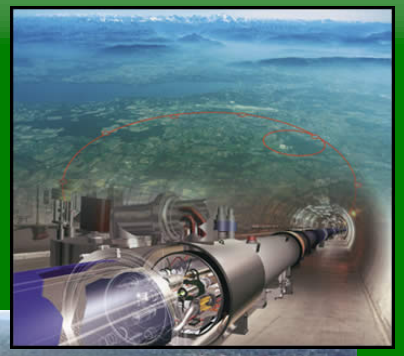


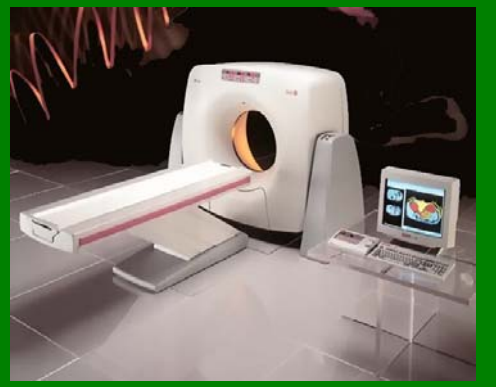




Elements of an e-Science Grid

- **Hardware**
  - Measuring devices
    - **Satellites in Astrophysics**
    - **Environmental Sensors**
    - **Scanners in medical imaging**
    - **Radiotelescopes in Astronomy**
    - **Detectors in particle physics accelarators**

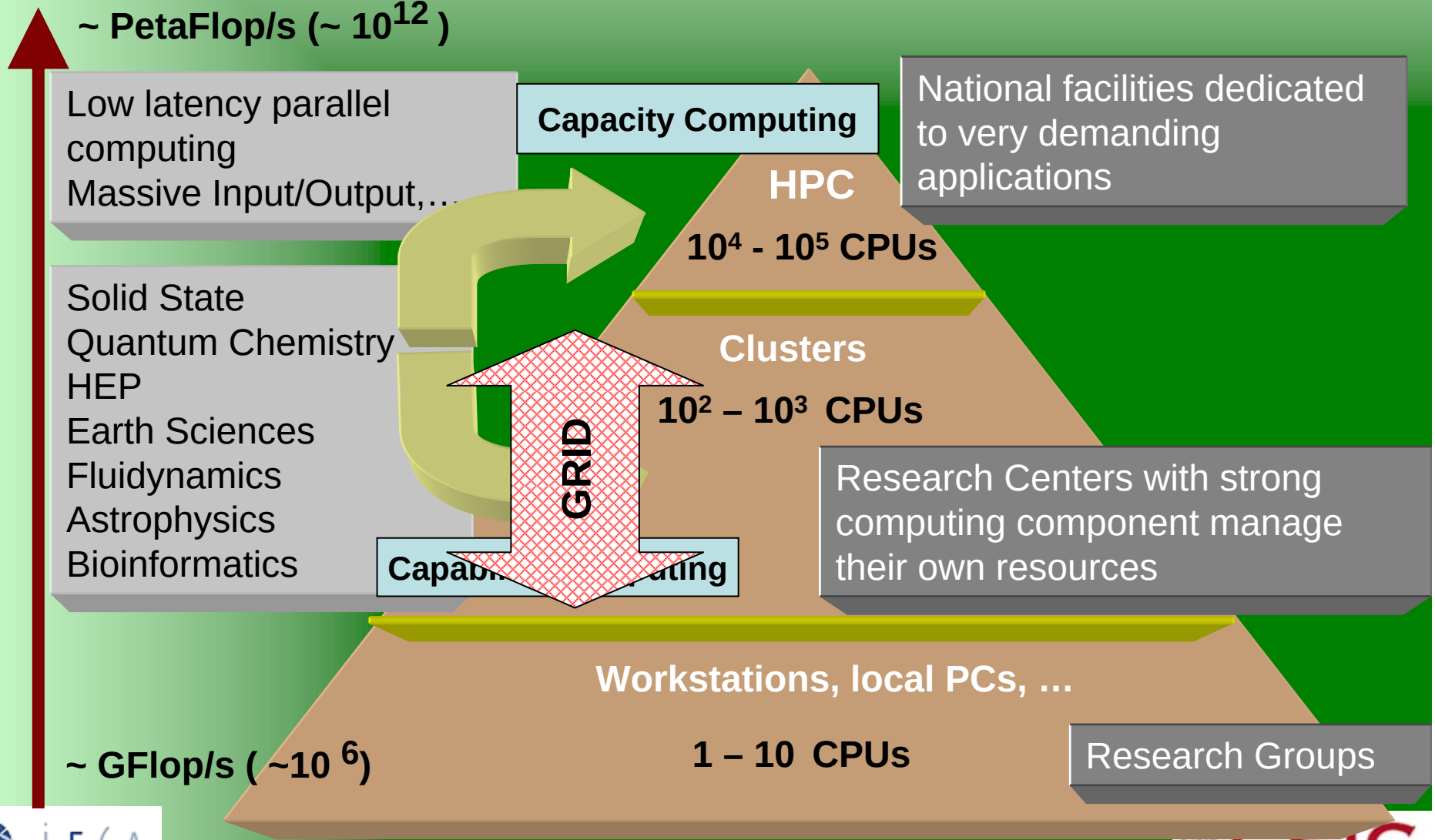


Elements of an e-Science Grid

- **Hardware**
  - Informatics
    - **DATA triggers**
    - **Storage Units (tapes, disks, etc…)**
    - **Computing facilities**

The Computing Piramide

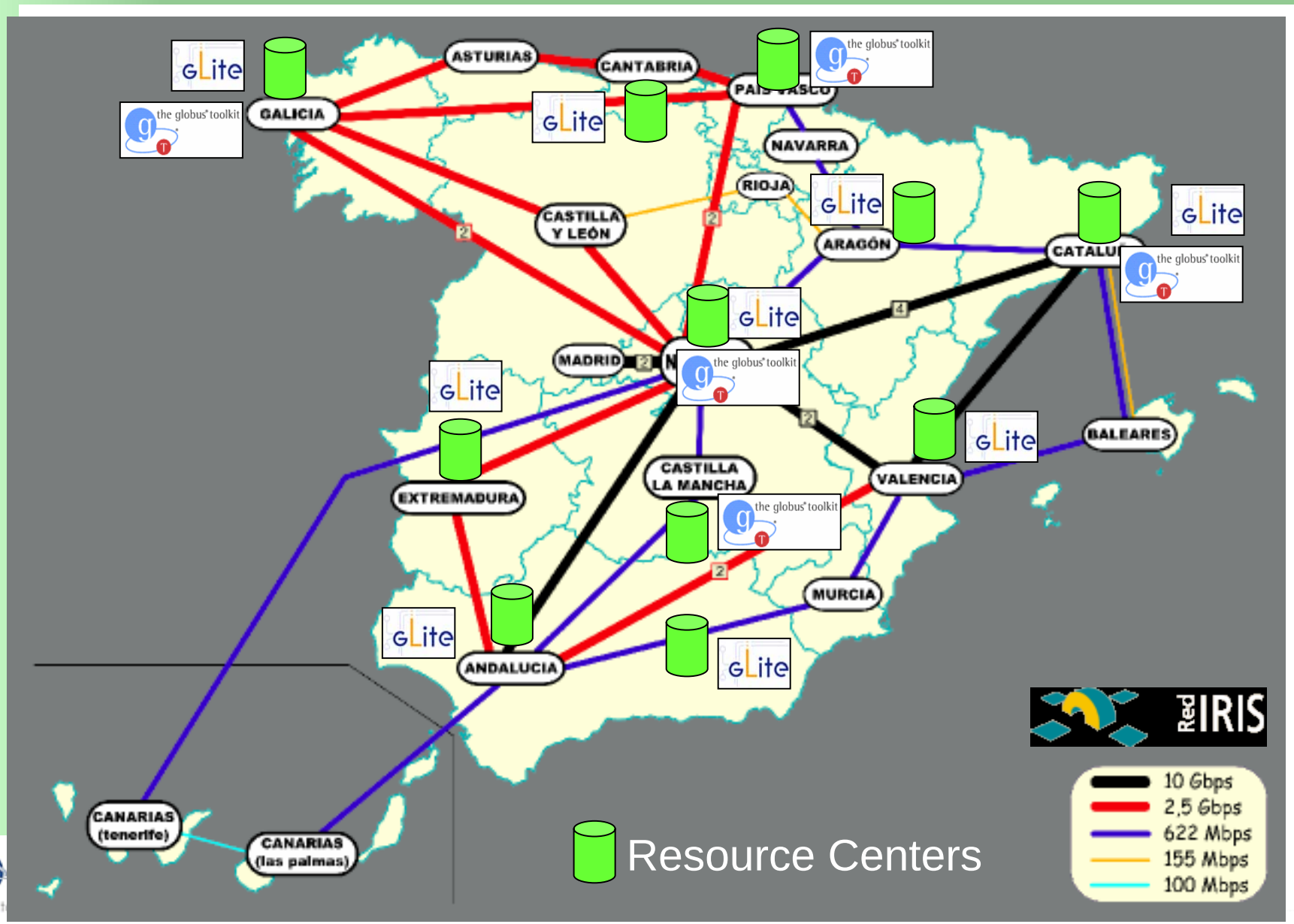


The Spanish Supercomputing Network (RES)

La UC es miembro de la Red Española de Supercomputación

- **El IFCA alberga el Nodo de la Univ. de Cantabria: *Altamira***
  - 256 Blades (512 PowerPC 2.2GHz)
  - Interconectadas con Myrinet 2000
  - Rendimiento 4.5 TeraFlop/s

The National Grid Infrastructure



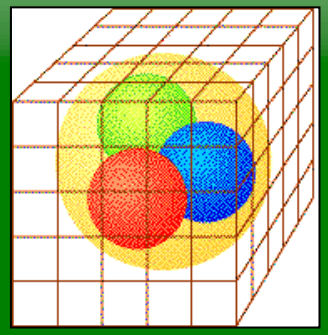
Elements of an e-Science Grid

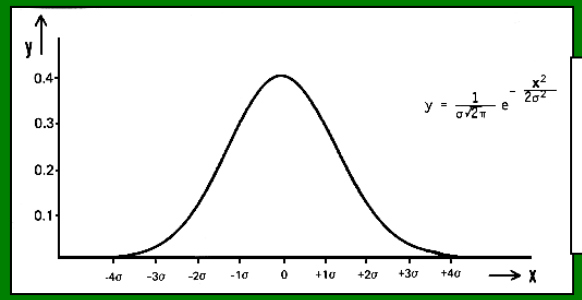
- **Software**
  - **Sistema Operativo compatible con el trabajo en Grid**
  - **Software intermediario (middleware) gestión:**
    - **Usuarios**
    - **Trabajos, datos,...**
    - **Infraestructura** de una forma transparente y segura
      - **Monitorización**
      - **Sistemas de Información** fiables
      - ....

Elements of an e-Science Grid

Software

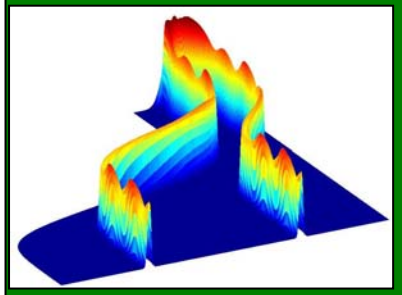
- **Aplicaciones**
  - Física Experimentall:
    - Altas Energías, Astrofísica
  - Biología Estructural
  - Física del Estado Sólido
  - Fusión Nuclear
  - Química Cuántica
  - Geofísica
- **Modo de trabajo**
  - Procesado de datos experimentales
  - Simulaciones de Monte Carlo
  - Dinámica y Mecánica molecular
  - Problemas de Fluidodinámica
  - Desarrollo de modelos predictivos





1. $(\frac{d^3y}{dx^3})^4 + 2\frac{dy}{dx} = \sin x$
2. $\frac{dy}{dx} - 2xy = x^2 - x$
3. $\frac{dy}{dx} - \sin y = -x$
4. $\frac{d^2y}{dx} = 2xy$

www.analyzemath.com

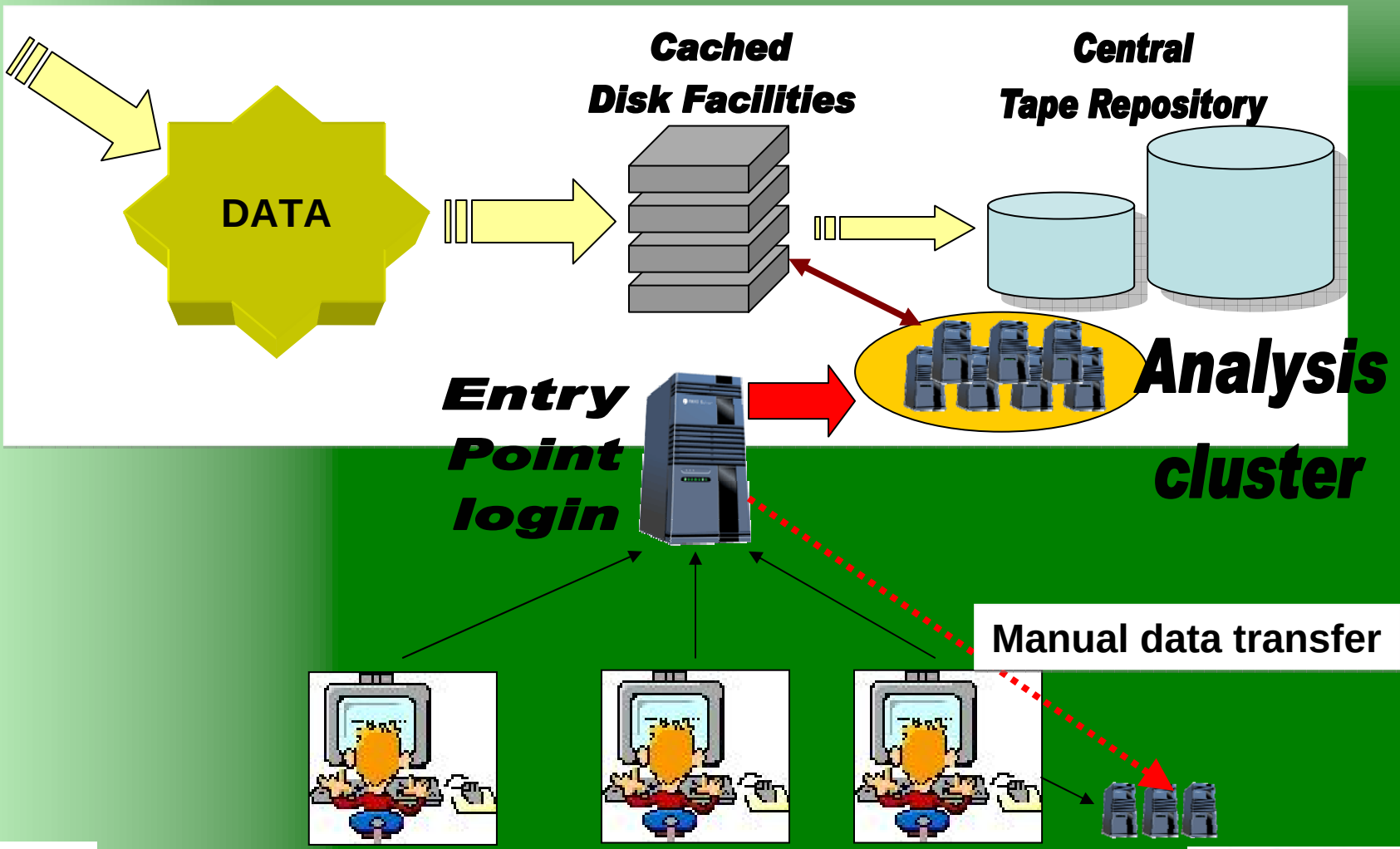


High Energy Physics: a Grid community

Key features

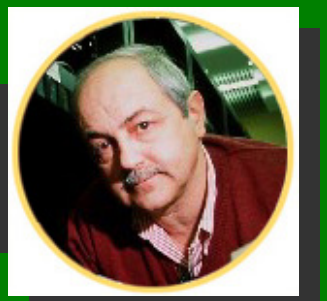
- Many (hundreds) research groups geografically dispersed
  - 20 years working in “Grid” but without Grid technology
- Analyzing large data sets from experiments in particle accelerators
- A very motivated community due to the success of the *World Wide Web*

Data Analysis in High Energy Physics



Deployment of Computational Grids
Building a Grid for HEP

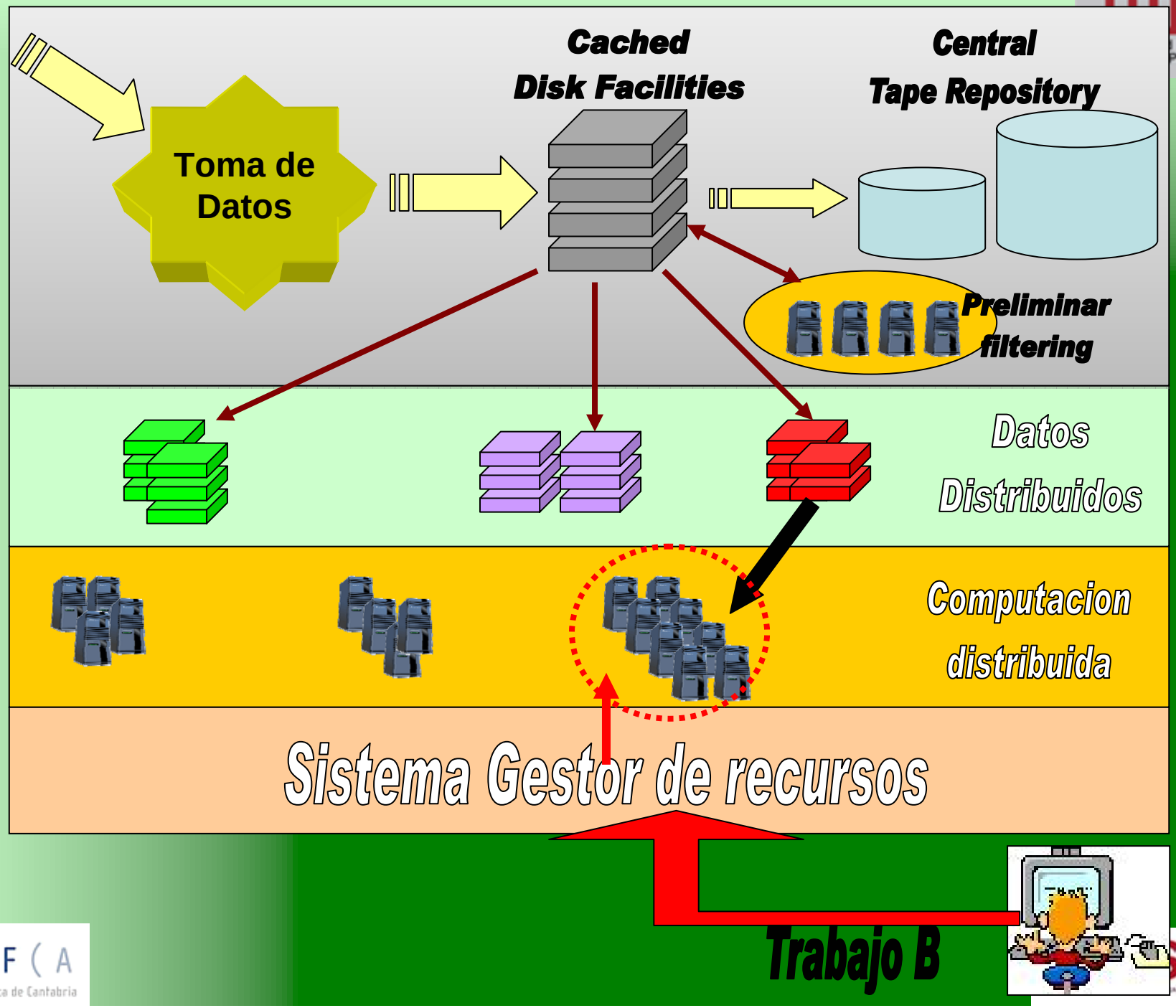
- Using as a basis the Globus Toolkit (gsi security, gridftp for data transfer, globus-job-run for job submission,…)
  - DATAGRID and its "succesor" EGEE develop a Grid middleware adapted to the needs of the experimental analysis in HEP
  - This middleware is called gLite

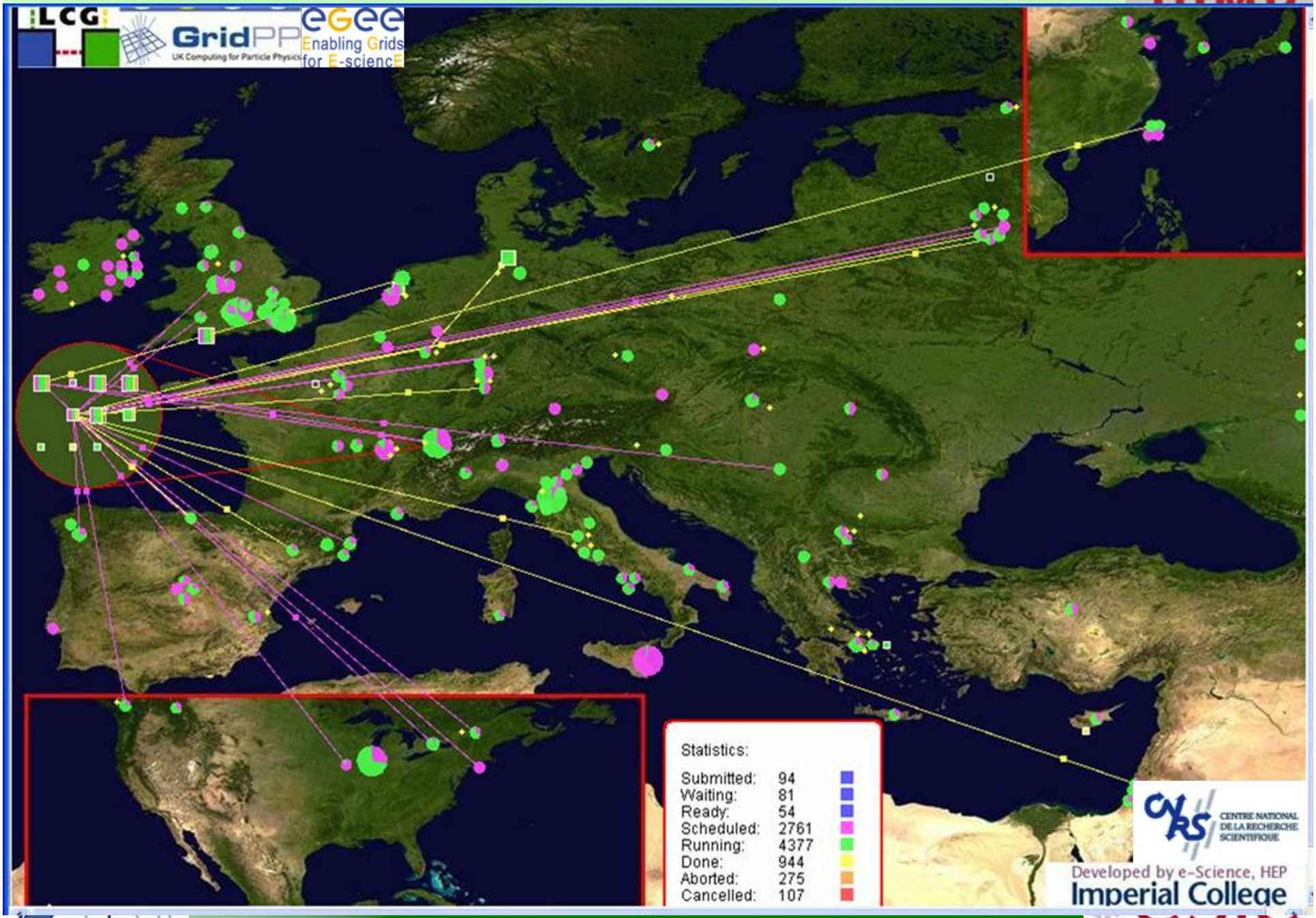


The current EGEE Grid is the largest distributed computing infrastructure

CPUs: ~ 40.000
Storage: ~ 5PBytes

and will provide the resources to analyze the data coming from the LHC detectors


Cached
Disk Facilities
Central
Tape Repository
Toma de
Datos
Preliminar
filtering
Datos
Distribuidos
Computacion
distribuida
Sistema Gestor de recursos
Trabajo B


LCG
GridPP
UK Computing for Particle Physics
eGee
Enabling Grids for E-sciencE
Statistics:
Submitted: 94
Waiting: 81
Ready: 54
Scheduled: 2761
Running: 4377
Done: 944
Aborted: 275
Cancelled: 107
CNRS
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
Developed by e-Science, HEP
Imperial College

Deployment of Computational Grids for other applications I

- Applications whose *modus operandi* is analogous to the analysis of experimental particle physics fit well on EGEE
  - Data are taken in a some kind of detector, and they are afterwards distributed to the researchers for analysis work

At IFCA the Grid experience developed for HEP (Tier 2 of CMS) has been profited for the group of Astrophysics and Astronomy

- Cosmic Microwave Background
  - **Planck Mision**
    - Analysis and image filtering techniques
- X-Ray Astronomy
  - Analysis of data from the satellite **XMM-Newton**

Deployment of Computational Grids for other applications II

Medio Ambiente

- **Computación Científica**
  - Modelización del clima, evolución de la atmósfera,
- **Monitorización de parámetros medioambientales**
  - Redes de sensores, European Global Monitoring Environment

Ciencia de Materiales

Computación Científica

Materiales magnéticos, nanoestructuras, mecánica de fluídos, cálculos de deformaciones con implicaciones en ingenierías,…

Astronomía, Astrofísica, Física Nuclear y de Partículas

Física Nuclear y de Partículas: LHC, ILC

Astronomía: Radio Astronomía, tratamiento de señales obtenidas en nuevas generaciones de telescopios

Astrofísica: Misión Planck

Energía

- **Fusión Nuclear: ITER**

Biomedicina y Ciencias de la Vida

Hemos asistido al surgimiento de la Biología como una Ciencia Computacional

- Secuenciado de ADN, análisis estructural de proteínas, visualización de alta resolución, captura de datos,…
- La bioinformática es un requisito para la biología (diseño de fármacos, genética, epidemiología, biología estructural)

Porting Applications to the Grid

- **Understanding the Application**
  - Description in terms of
    - Area of knowledge and status of the art
    - Results expected and impact on the scientific community
    - Understanding the computational approach at the algorithmic level
  - Resources needed
    - Software & Hardware
    - GRID services
  - GRID added value
    - Why on the GRID ?
- **Interactive/inmediate access to resources**
- **Graphics & Visualization**
- **Quality of Service regarding network reliability**

Middleware: shotcomings of EGEE and gLite

- Applications on EGEE
  - EGEE was built thinking on HEP applications
    - The four experiments of LHC
    - Optimized for Monte Carlo runs of small duration (few hours)
- Generic Applications are more complicated to support
  - Longer runs
  - Dependencies on external software and libraries
  - Specific porting issues
  - **Supportin MPI parallel jobs anf interactive access requires going beyond gLite**

Program of the week

- **Grid Applications areas**
  - Biomedicine *(Monday)*
  - Astrophysics *(Tuesday)*
  - Nuclear Fusion *(Wednesday)*
  - High Energy Physics *(Thursday)*
- **Grid middleware**
  - Introduction to glite (now)
  - Advanced middleware (*Tuesday, Wednesday morning*)
  - Virtualization (Tuesday morning)
  - BOINC schedulers (Wednesday afternoon)
  - Tutorial on virtualization (Thursday afternoon)
- **Evolution of Grid infrastructures in Spain and Europe**
  - Spanish NGI, European Grid Initiative (Friday morning)

Subject	Monday	Tuesday	Wednesday	Thursday	Friday
Biomedicine	X				
Astrophysics		X			
Nuclear Fusion			X		
HEP				X	
Middleware	X	X	X		
Infrastructure			X		
Grids of Instrumentation				X	
Virtualization		X		X	
Future					X

Alumnos del Master de Computación

- Entregar un trabajo resumen del curso que contemple los aspectos generales (etre 5 y 10 paginas mínimo a 12pt)
  - ¿Qué es el Grid?
  - Aplicaciones científicas en el Grid
- Posibles Aplicaciones de la tecnología Grid en su área de trabajo o de investigación
- Plazo: hasta el 20 de Julio de 2009 (entregar en la secretaría del IFCA)