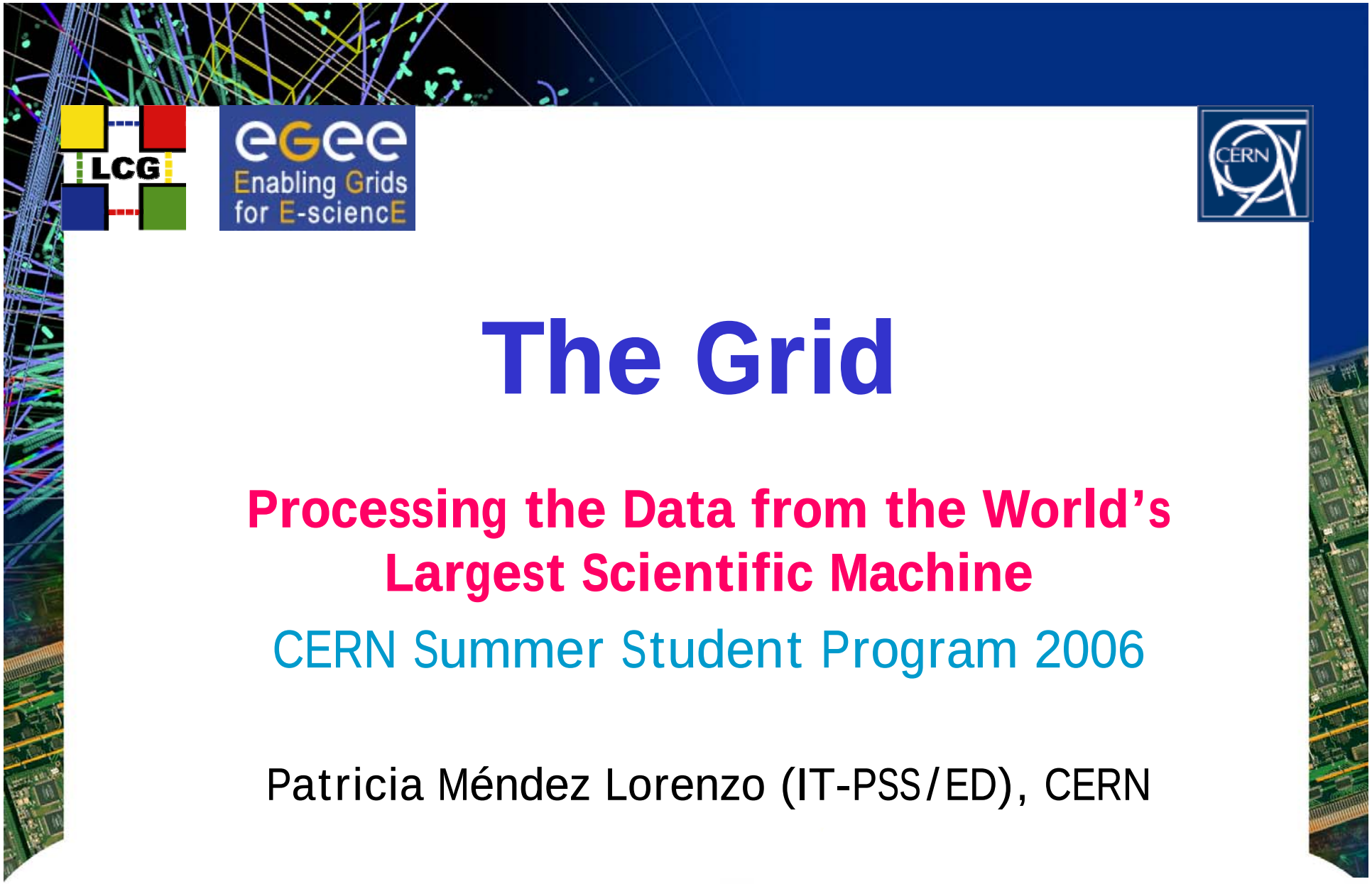


The Grid

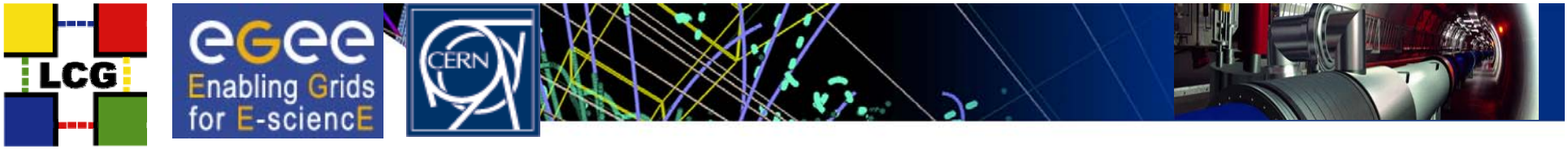
Processing the Data from the World's Largest Scientific Machine

CERN Summer Student Program 2006

Patricia Méndez Lorenzo (IT-PSS/ED), CERN

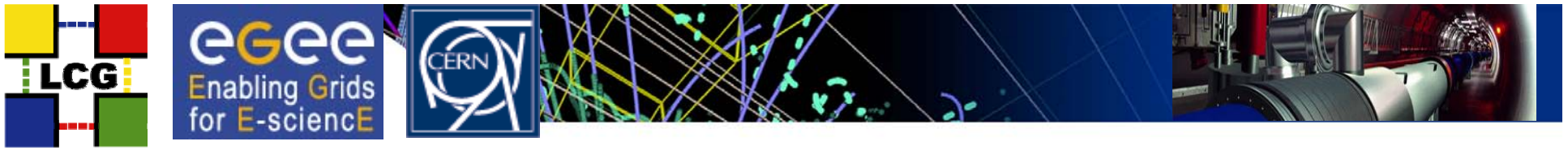
Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing



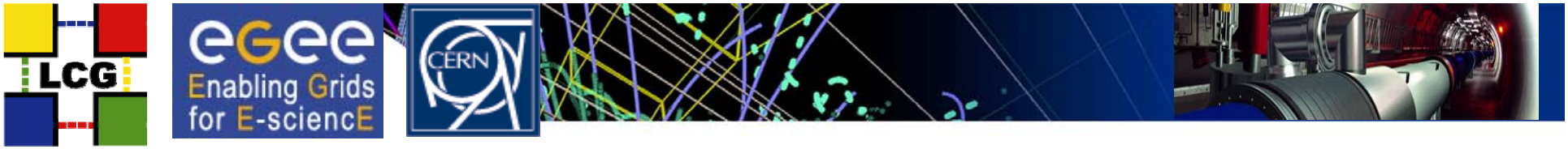
Abstract

- The world's **largest scientific machine** will enter production about one year from the time of this talk
- In order to exploit its scientific potential, computational resources way beyond those needed for previous accelerators are required
- Based on these requirements, a distributed solution based on **Grid technologies** has been developed
- This talk describes the overall **Grid infrastructure** offered to the experiments, the state of **deployment** of the production services, and the applications of the Grid beyond the HEP



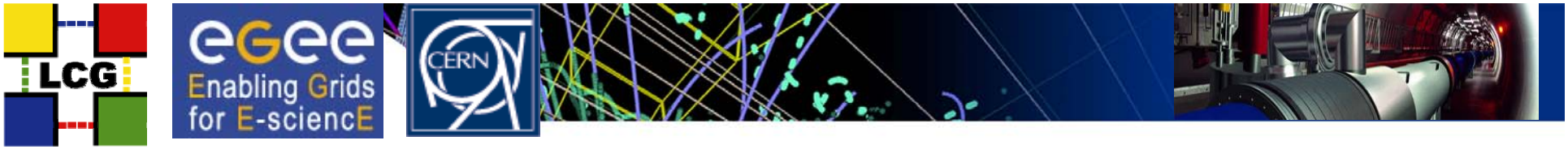
Several considerations before beginning…

- I assume that you have never submitted a job to the Grid
  - No large knowledge of computing is required
- **Announcement to the Physicist!**
  - "…This is a computer related talk, no interest for me…" **(WRONG!)**
  - If you do not know the software infrastructure of your experiment and you do not know computing, you will have big problems in your future work as researcher
- This presentation does not pretend to be a tutorial
  - I will avoid as much as possible the use of code
- This presentation will show you the infrastructure that many of you will have to face as researchers and it begins to be expanded in many fields of research (not only HEP)



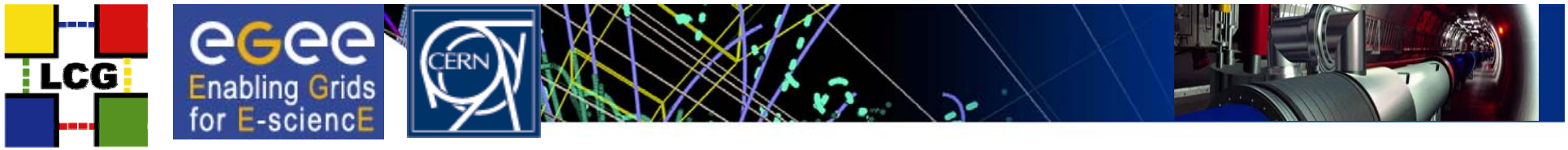
Overview

- First Part: The general concept of the Grid
- Second Part: The WLCG Grid Computing
- Third Part: The Infrastructure and the Services
- Fourth Part: Some applications
- Fifth Part: Current Status and applications beyond HEP



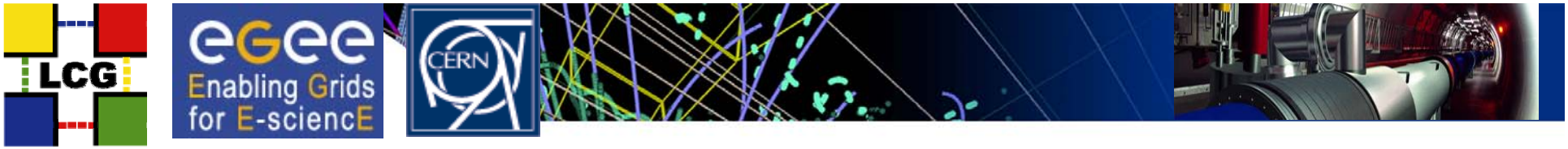
1st Part

- General Concept of the GRID



1st Part: What is a Grid

- A genuine new concept in distributed computing
  - Could bring radical changes in the way people do computing
  - **Share** the computing power of many countries for your needs
  - **Descentralized** the placement of the computing resources
- A hype that many are willing to be involved in
  - Many researchers/companies work on "grids"
    - More than 100 projects (some of them commercial)
    - Only few large scale deployments aimed at production
  - Very confusing (hundreds of projects named Grid something)
- Interesting links: Ian Foster, Karl Kesselman
  - Or have a look at http://globus.org
    - Not the only grid toolkit, but one of the first and most widely used
  - Web pages of: EGEE, WLCG (we will talk about them)



What is a Grid (cont.)

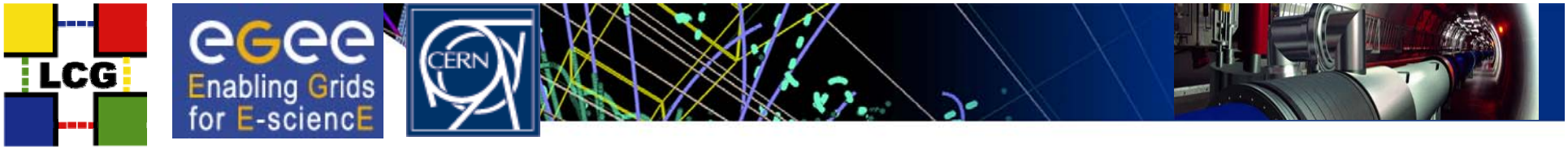
- Basic concept is simple:
  - I.Foster: "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. "
  - Or:"On-demand, ubiquitous access to computing, data, and services"
  - From the user's perspective:
    - I want to be able to use computing resources as I need
    - I don't care who owns resources, or where they are
      - Have to be secure
      - My programs have to run there
  - The owners of computing resources (CPU cycles, storage, bandwidth)
    - My resources can be used by any authorized person (not for free)
    - Authorization is not tied to my administrative organization
  - **NO centralized control of resources or users**

Basic Philosophy

Just one idea: How to see the Grid

- The Grid connects Instruments, Computer Centers and Scientists

- If the WEB is able to share information, **the Grid is intended to share computing power and storage**

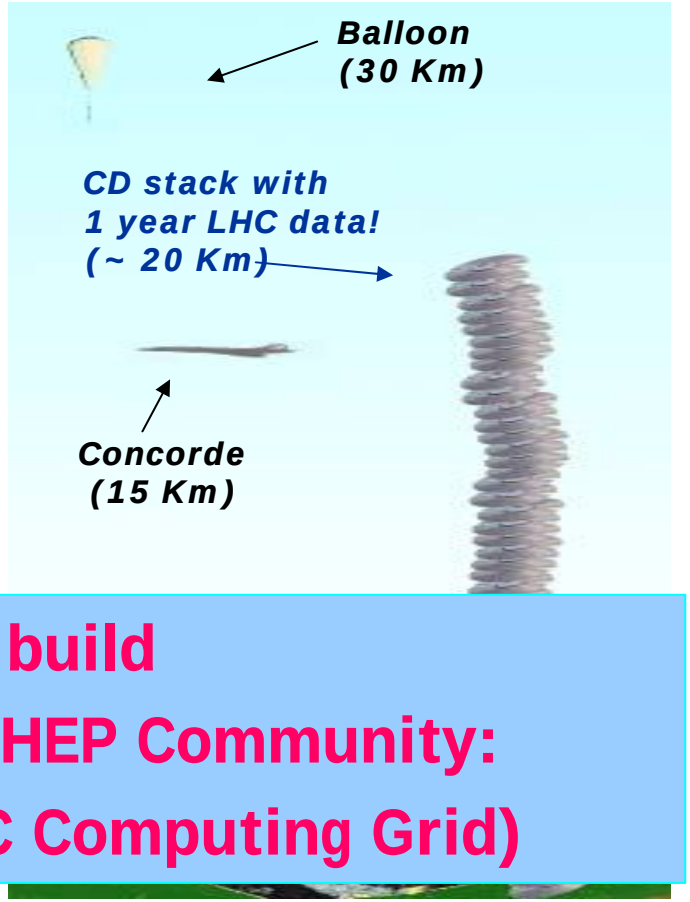


2nd Part

- The WLCG GRID Computing

The LHC Grid Computing

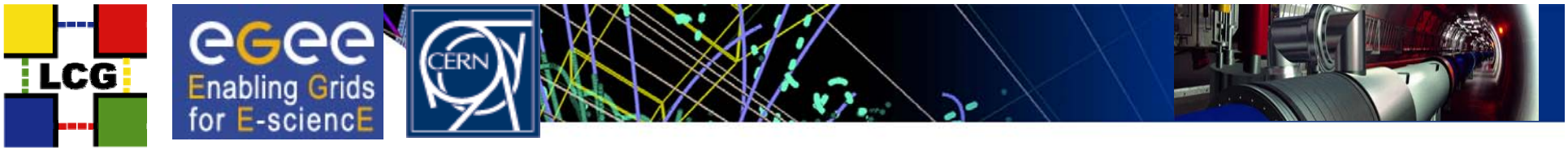
- LHC will be completed in 2007 and run for the next 10-15 years
- Experiments will produce about **15 Millions Gigabytes per year** of data (about 20 million CDs!)
- LHC data analysis requires a computing power equivalent to around **100000 of today's fastest PC processors**
- R
  c
  o
  c



Therefore we build

a Computing Grid for the HEP Community:

The WLCG (Worldwide LHC Computing Grid)



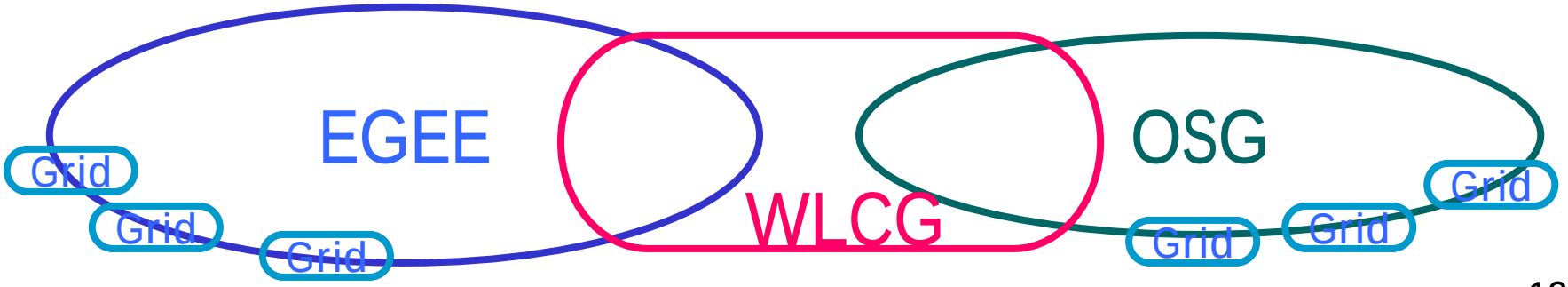
More reasons to use the GRID

The LCG Technical Design Report lists:

1. Significant costs of maintaining and upgrading the necessary resources … more easily handled in a distributed environment, where individual institutes and organizations can fund local resources whilst contributing to the global goal
2. … no single points of failure. Multiple copies of the data, automatic reassigning of tasks to resources … facilities access to data for all scientists independent of location. … round the clock monitoring and support.

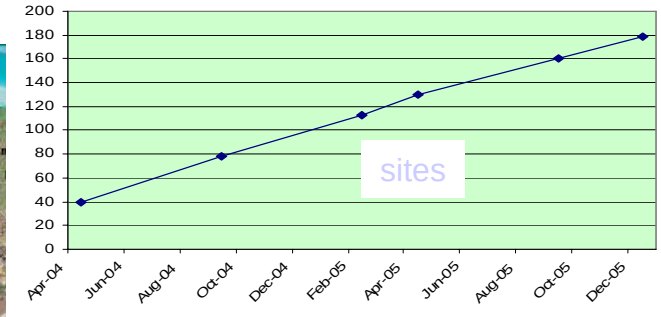
Projects we use for LHC Grid

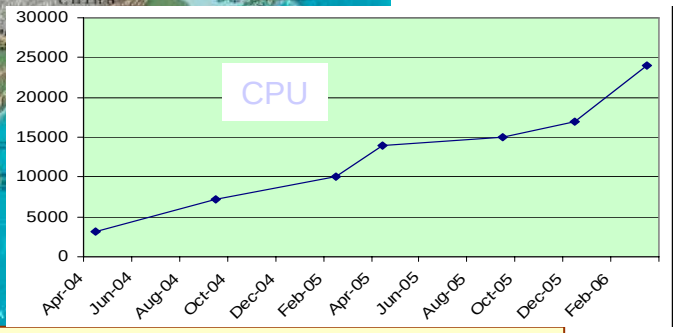
- WLCG depends on 2 major science grid infrastructures provided by:
  - EGEE – Enabling Grid for E-SciencE (160 communities)
  - OSG – US Open Science Grid (also supporting other communities beyond HEP)
  - Both infrastructures are federations of independent GRIDs
  - WLCG uses both OSG resources and many (but not all) from EGEE



EGEE Sites

EGEE:
Steady growth over the lifetime of the project





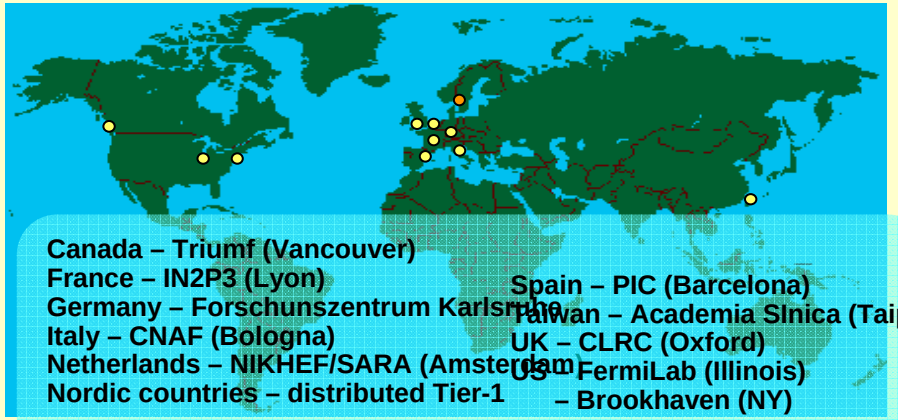
EGEE:
> 180 sites, 40 countries
> 24,000 processors,
~ 5 PB storage

country	sites	country	sites	country	sites
Austria	2	India	2	Russia	12
Belgium	3	Ireland	15	Serbia	1
Bulgaria	4	Israel	3	Singapore	1
Canada	7	Italy	25	Slovakia	4
China	3	Japan	1	Slovenia	1
Croatia	1	Korea	1	Spain	13
Cyprus	1	Netherlands	3	Sweden	4
Czech Republic	2	FYROM	1	Switzerland	1
Denmark	1	Pakistan	2	Taipei	4
France	8	Poland	5	Turkey	1
Germany	10	Portugal	1	UK	22
Greece	6	Puerto Rico	1	USA	4
Hungary	1	Romania	1	CERN	1

The WLCG Distribution of Resources

Tier-0 – the accelerator centre

- Data acquisition and initial Processing of raw data
- Distribution of data to the different Tier's



Tier-1 (11 centers) – "online" to the data acquisition process → high availability

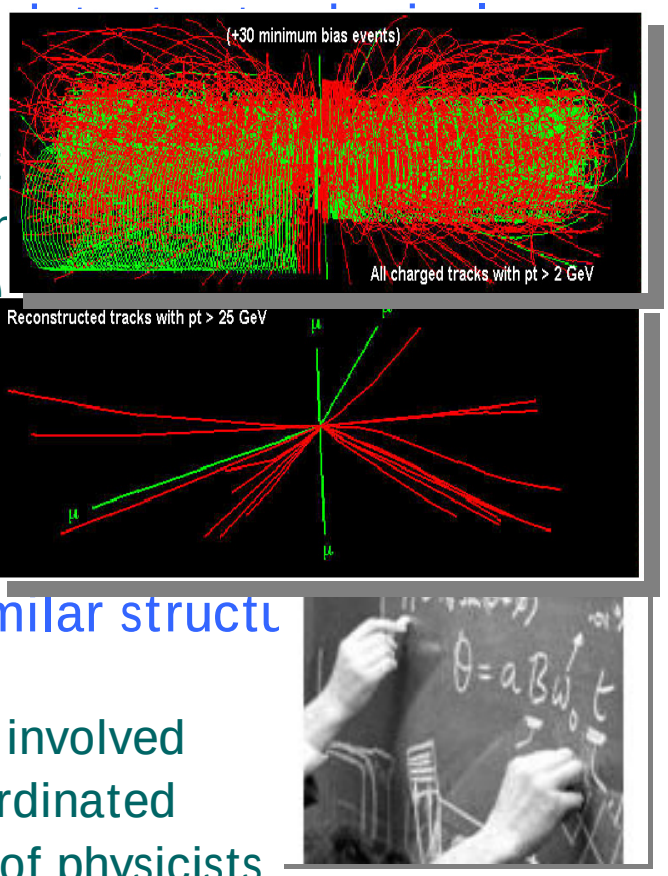
- Managed Mass Storage – → grid-enabled data service
- Data-heavy analysis
- National, regional support

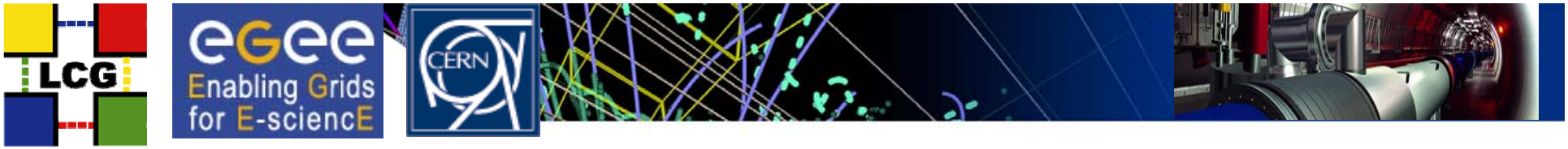
Tier-2 – ~100 centres in ~40 countries

- Simulation
- End-user analysis – batch and interactive

From the raw data… to the paper

- **Reconstruction**: transform signals from the
  - energy, charge, tracks, momentum, part
  - this task is computational intensive and h
  - structured activity (production manager)
- **Simulation**: start from the theory and comp detector
  - very computational intensive
  - structured activity, but larger number o
- **Analysis**: complex algorithms, search for similar structu physics
  - very I/O intensive, large number of files involved
  - access to data cannot be effectively coordinated
  - iterative, parallel activities of hundreds of physicists

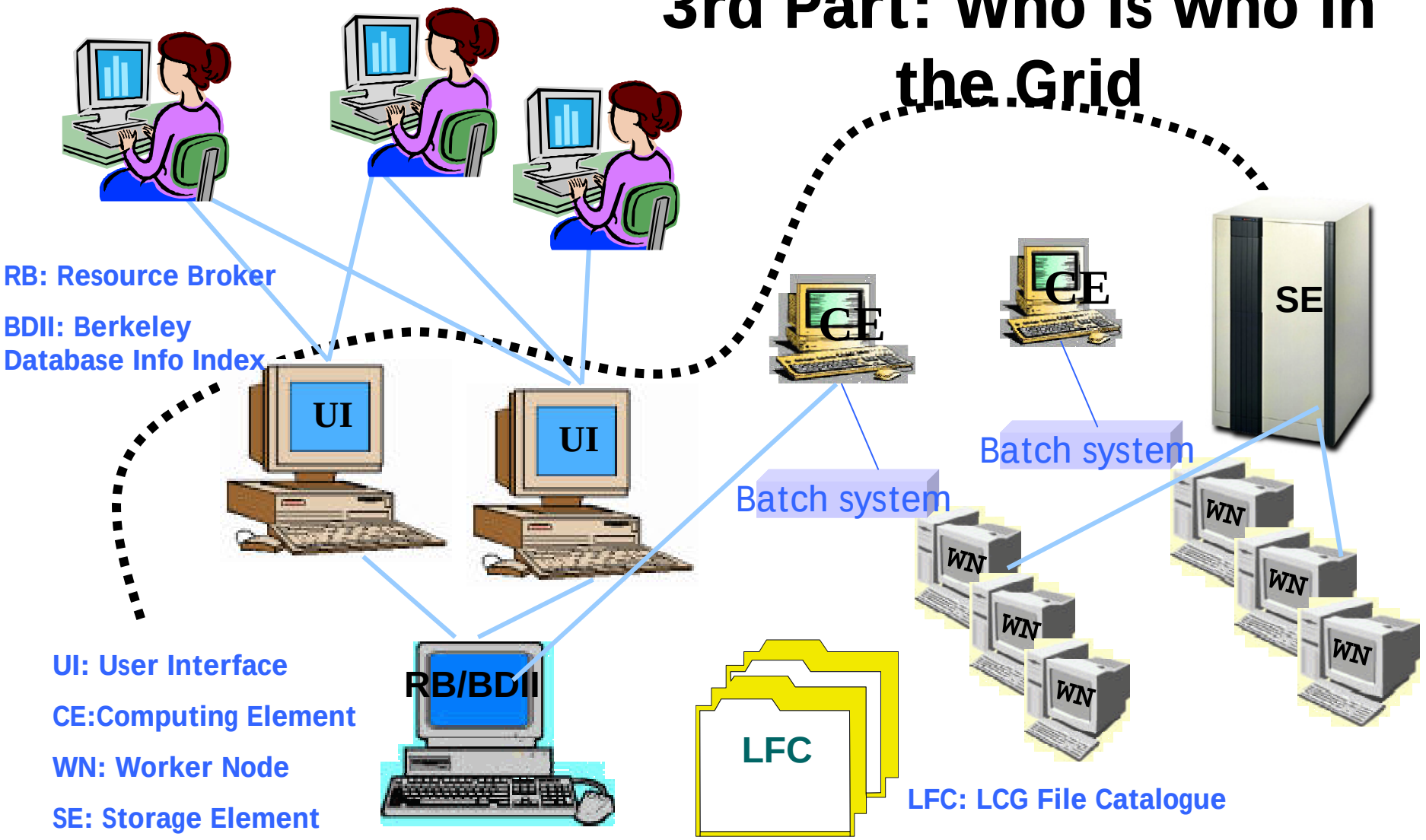




3rd Part

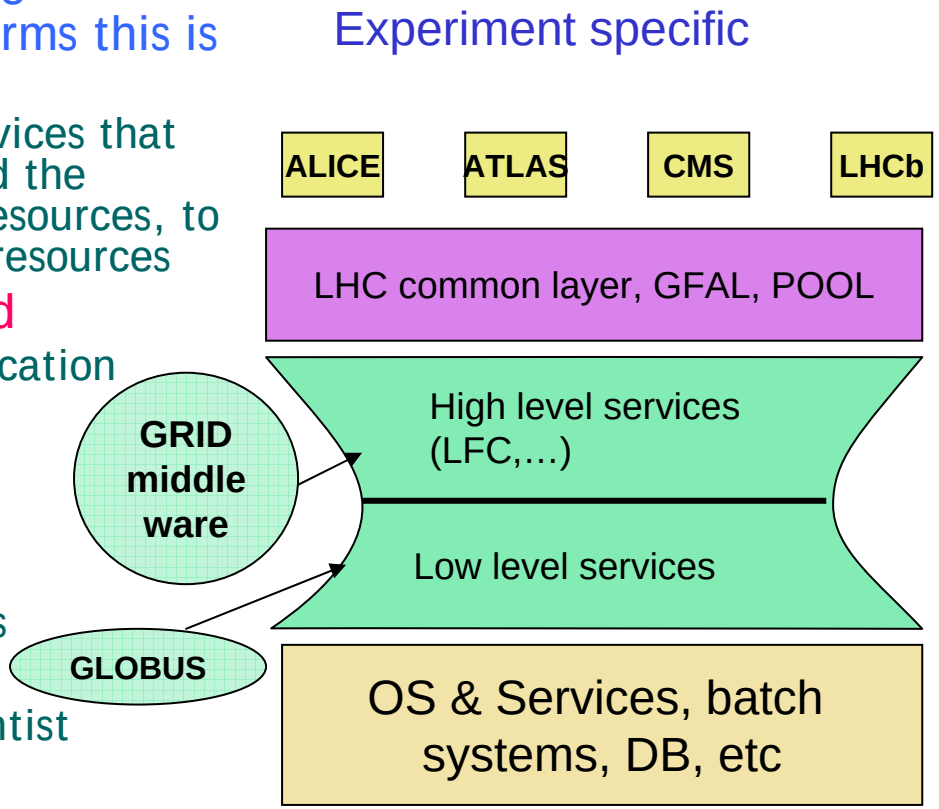
- The Infrastructure and the Services

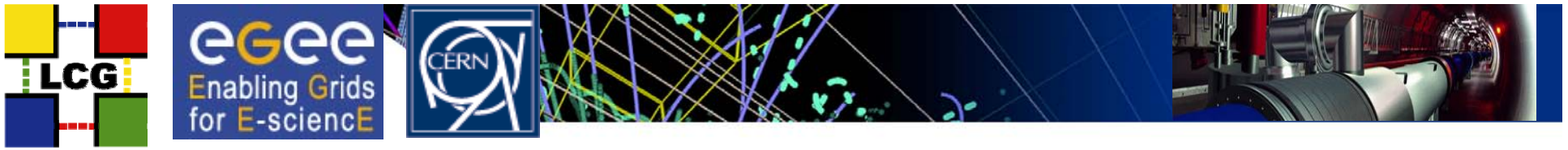
3rd Part: Who is who in the Grid



Middleware: what it is

- All these services and their tools should talk the same language to make the whole mechanism able to work … in Grid terms this is called: MIDDLEWARE
  - The middleware is software and services that sit between the user application and the underlying computing and storage resources, to provide an uniform access to those resources
- The Grid middleware services: should
  - Find convenient places for the application to be run
  - Optimize the use of the resources
  - Organize efficient access to data
  - Deal with security
  - Run the job and monitor its progress
  - Recover from problems
  - Transfer the result back to the scientist

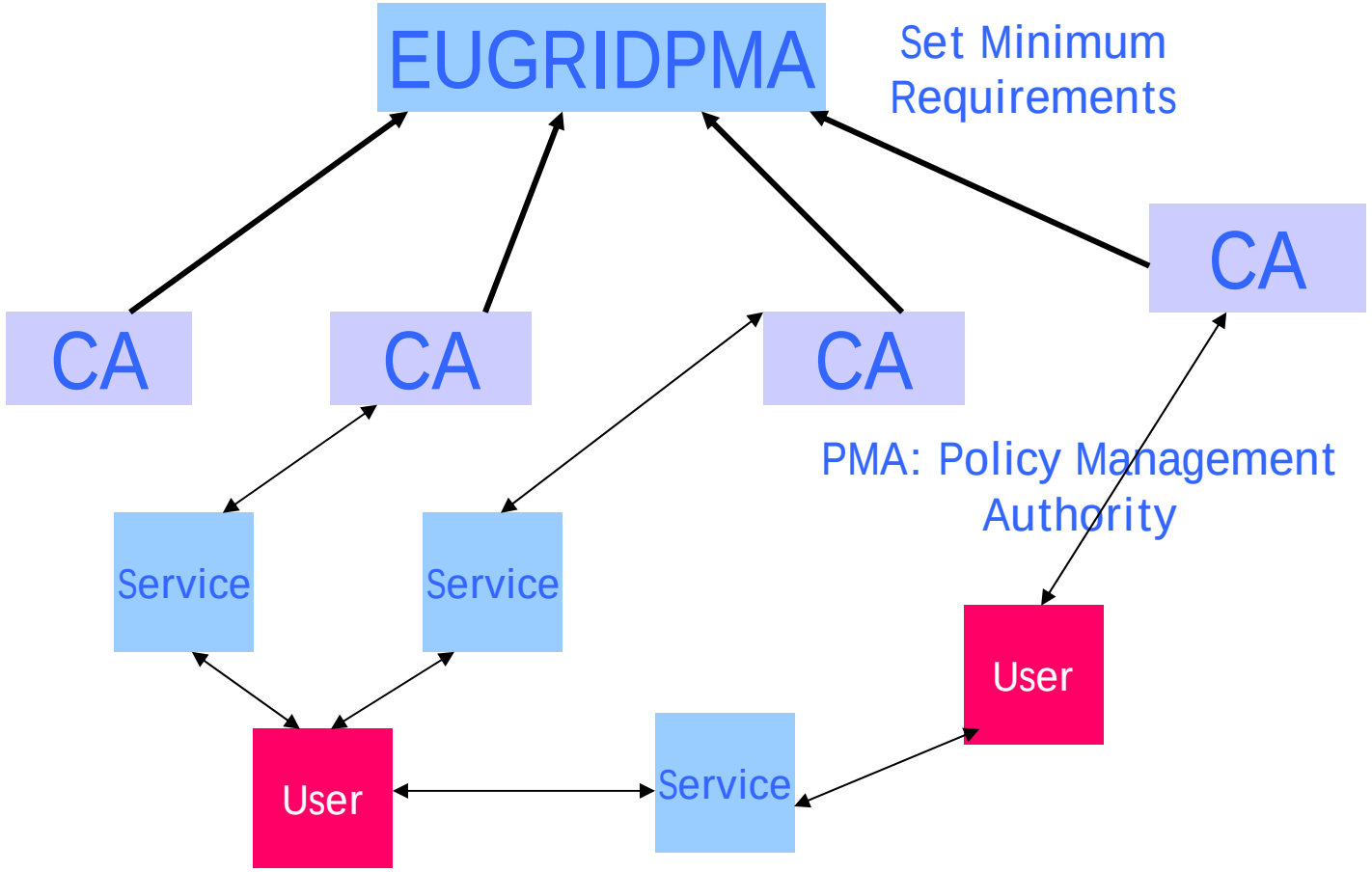


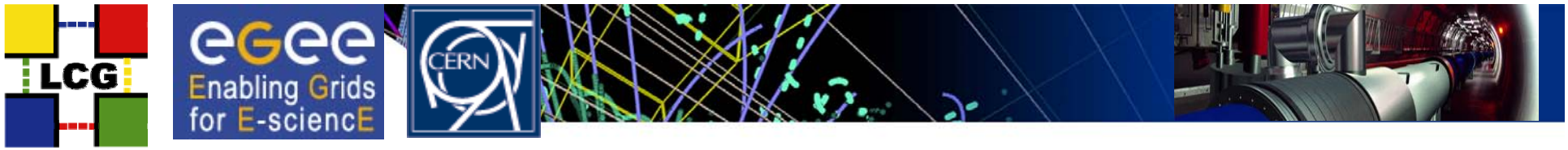


Fundamental: Security

- Security global mechanism
  - GSI: Globus Security Infrastructure
  - Based on PKI (public/private key infrastructure)
- Authentication: done via **CERTIFICATES**
  - To ensure that you are who you say to be
    - A certificate is a file which contains the public key, the lifetime of the certificate and the Distinguish Name (DN) of the user
    - This is signed by a CA which provides you the certificate
- At the core a set of Certification Authorities (CA)
  - Issue certificates for users and services authentication (valid 1 year)
  - Revokes certificates if required
  - Publish formal document about how they operate the service
  - Each site can decide which CA they accept and which ones they deny
- Authorization
  - Determines what you can do

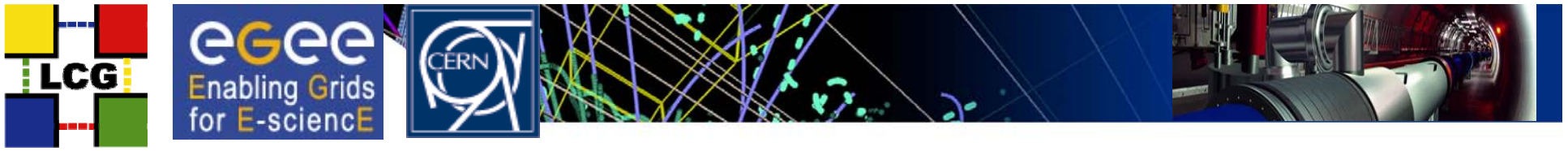
How the security works in the Grid (cont.)





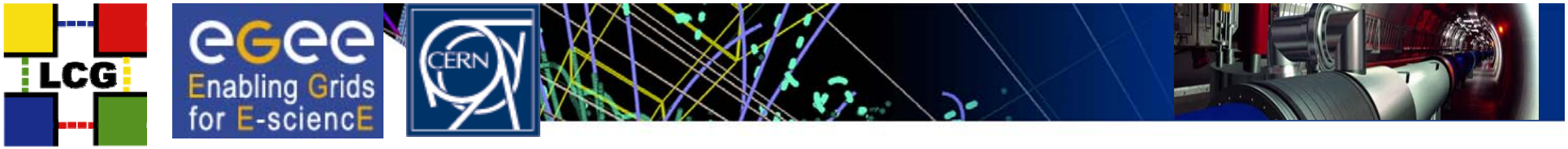
The proxy

- This is very nice but it has a nasty problem
  - The message is encrypted with the private key
  - This private key is encrypted in my local area with a password
  - The an encryption cannot be done with an encrypted private key
- Solution: The proxy
  - The proxy is a certificate
  - But contains the private key no encrypted!
  - So be careful!! You should not create proxies of very long life time



Authorization

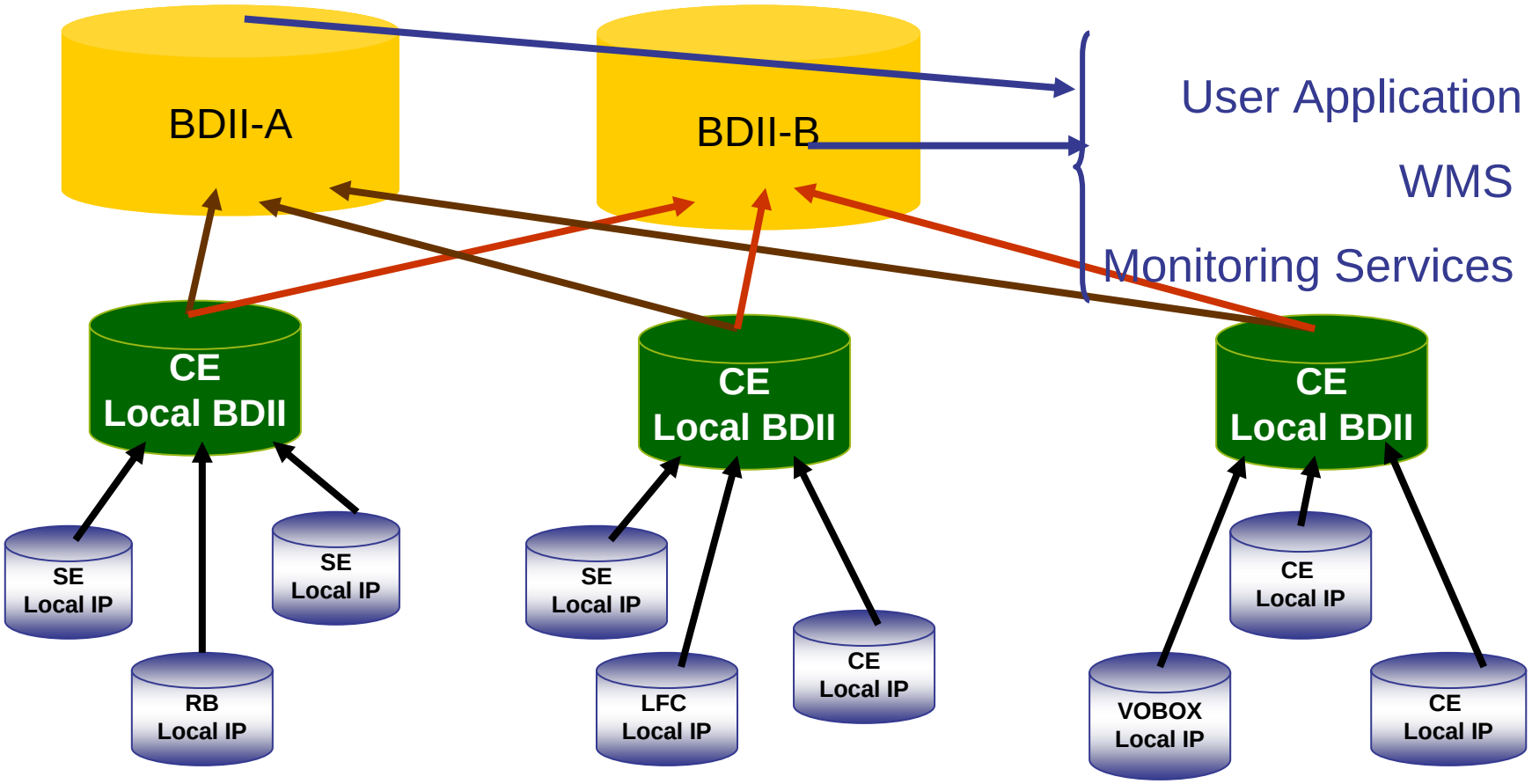
- Virtual Organizations (VO)
  - Homogeneous groups of people with common purposes and Grid rights
  - WLCG VO names: atlas, cms, alice, lhcb, dteam
  - At this moment, while registering with a VO, you get authorized to access the resources that the VO is entitled to use
- Site providing resources
  - Decides which VO will use its resources (negotiation with the experiments)
  - Can block individual persons of using the site
- VOMS (VO Membership Service)
  - System that clarifies users that are part of a VO on the base of attributes granted to them upon request
  - Allows definition of rules in a very flexible way
    - Production manager, software manager, simple user, etc.



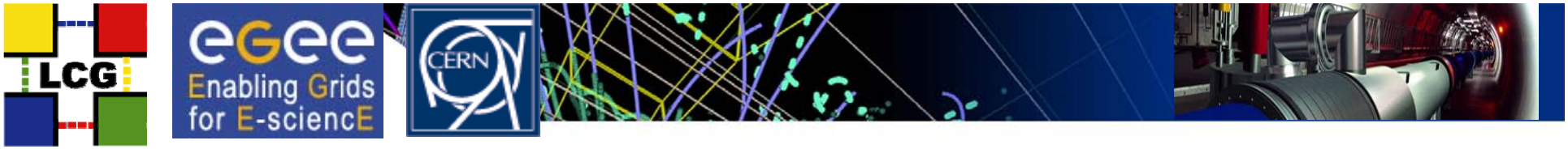
Information System

- The 1st thing that you need arriving in a new city, country…. is **INFORMATION**
  - Before using the system, see what the system can provide you
  - And information is given for free
    - No proxies, or certificates are needed
- The Information System of the LHC Grid Computing has a clear definition and it is hierarchically distributed

Collectors and Information Providers



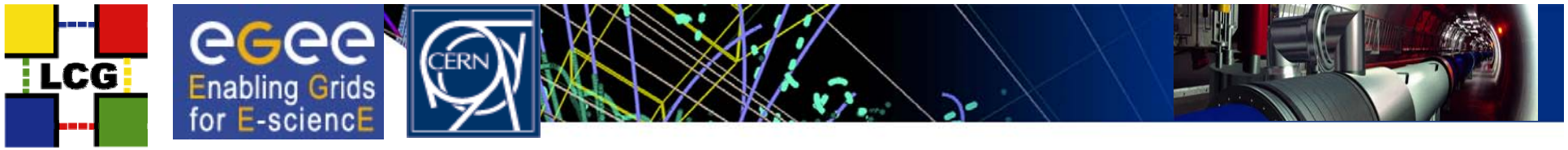
IP: Information Provider



Workload Management System

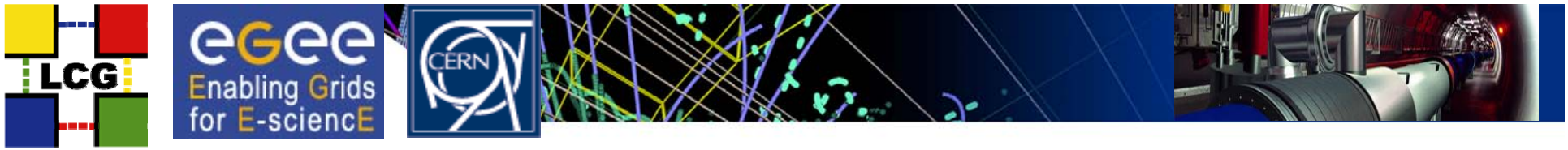
- This is the service that matches resources with jobs
  - Runs on a node called **RB** (Resource Broker)
  - Keeps track of the status of jobs (**LBS** Logging and Bookkeeping Service)
  - Talks to the batch systems on the remote sites (CE)
  - Matches jobs with sites where data and resources are available
  - Re-submission if jo
- The user describes
  (Job Description L

```
Executable = "gridTest";
StdError = "stderr.log";
StdOutput = "stdout.log";
InputSandbox = {"home/joda/test/gridTest"};
OutputSandbox = {"stderr.log", "stdout.log"};
Requirements =
  other.GlueHostOperatingSystemNameOpSys ==
  "LINUX"  && other.GlueCEStateFreeCPUs>=4;
```



Data Management System

- In the past distributed computing was quite focused on the computational aspect
  - Supporting large distributed computational tasks
  - Managing the sharing of the network bandwidth
- Today data access is becoming the main bottleneck
  - Huge amount of data (~PB)
  - Distributed in many sites
  - Not all of them can be replicated
- What a Storage Element should provide
  - The SE resources have to provide a good access and services to storage spaces
  - Middleware has to be able to manage different storage systems uniformly and transparently for the user

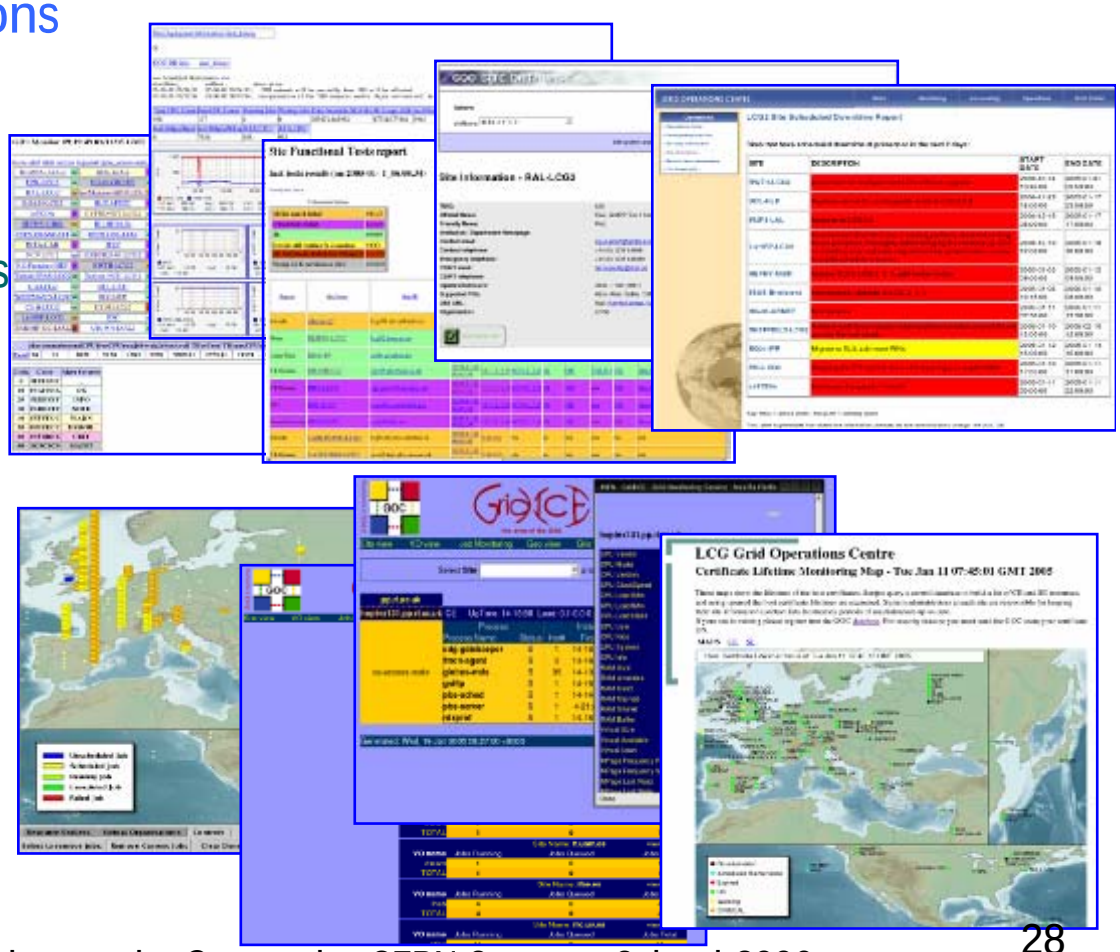


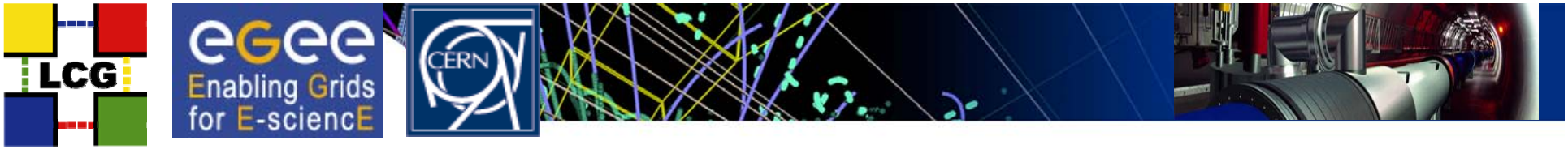
Data Management System

- SE supported by WLCG
  - Depending on the infrastructure of the site, different Storage system are supported by the WLCG
  - All of them with different implementations
  - Big and complex Storage system are normally owned by T1 sites
    - MSS (Mass Storage System as castor or dcache). Disk front ends with a tape backend
  - Disk arrays (DPM: Disk Pool Managers) have been developed as solution for T2 sites
- The SE are managed by the Storage Resource Manager (SRM)
  - It will hide the different Storage implementations
  - File pinning
  - Space reservation

Monitoring Systems

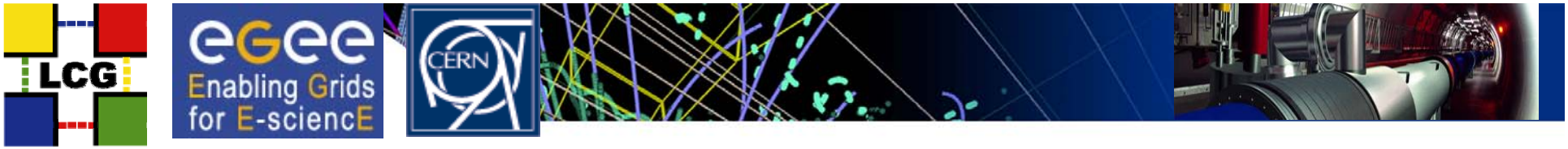
- Operation of Production Service: real-time display of grid operations
- Accounting information
- Selection of Monitoring tools:
  - GIIS Monitor + Monitor Graphs
  - Sites Functional Tests
  - GOC Data Base
  - Scheduled Downtimes
  - Live Job Monitor
  - GridIce – VO + fabric view
  - Certificate Lifetime Monitor





4th Part

- Some Applications

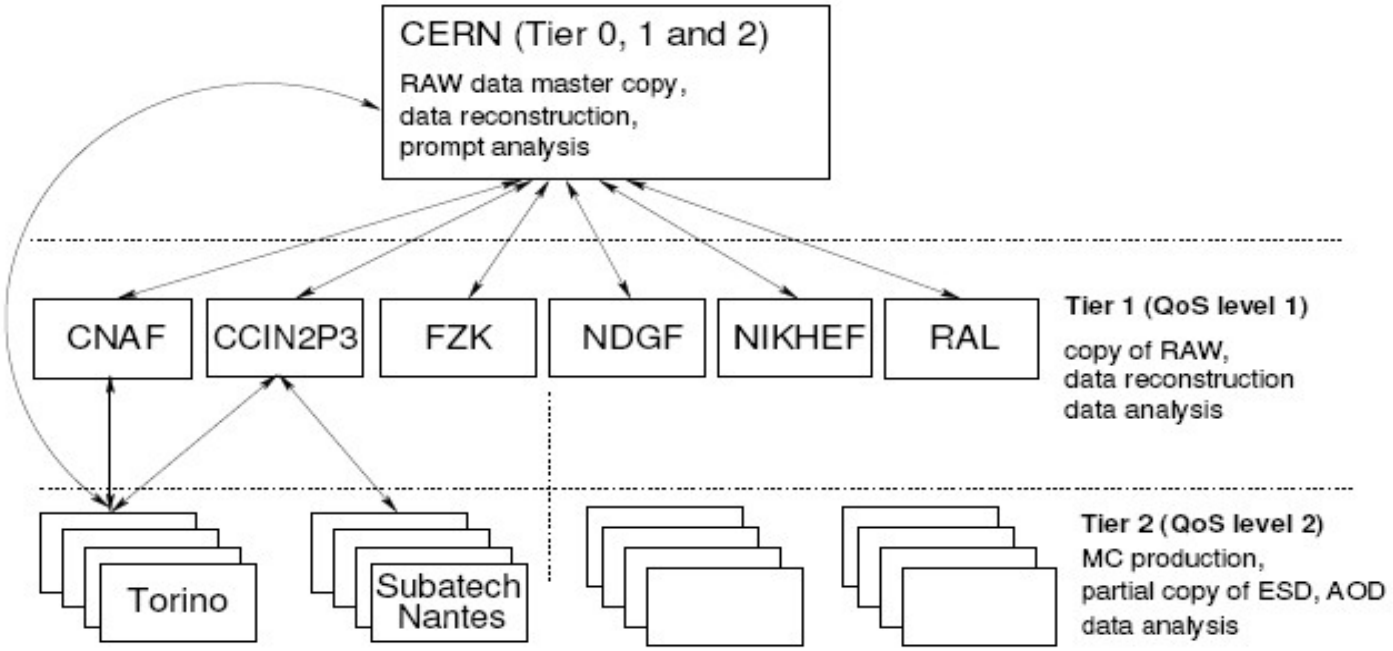


Services and Support

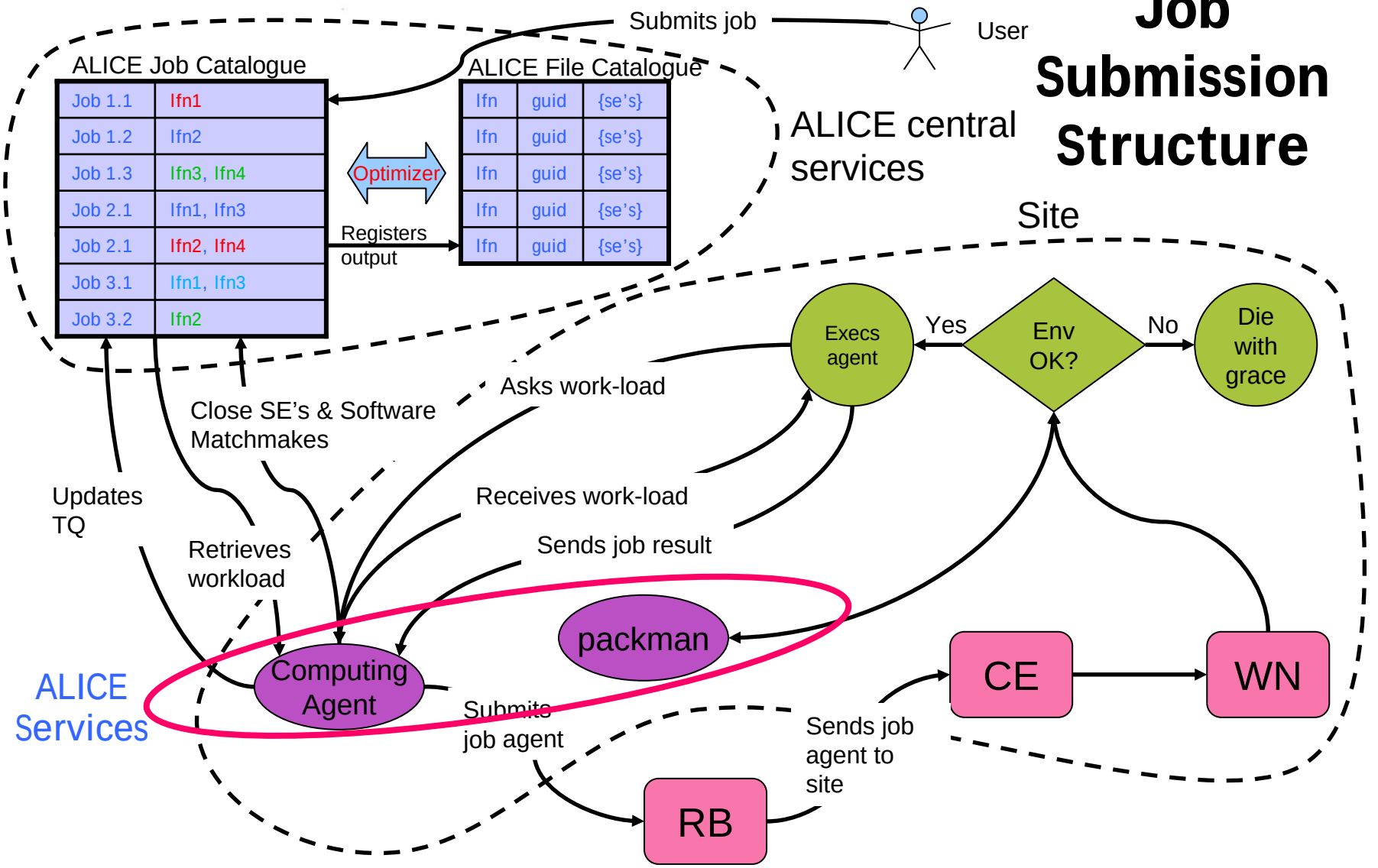
- All experiments have already decided the rule that each Tier will play in their production
- The **offline groups** of each experiment have to provide to their end-users with a stable, feasible and user friendly infrastructure to access the data, run their jobs and analyze the data
- The best way to get ready is "emulating" the real data taking performing the so-called: **Data Challenges**
  - **Validation of their computing systems**
  - **Validation and testing of the WLCG infrastructure**
- The Grid services are provided by the WLCG developers but always in communication with the experiments
  - The aim is to make a successful data taking, we always try to provide the experiments with what they require
- A complete infrastructure of support has been developed for each experiments
  - In this sense, each experiment has a WLCG specific contact person + contact with the sites + support of the services

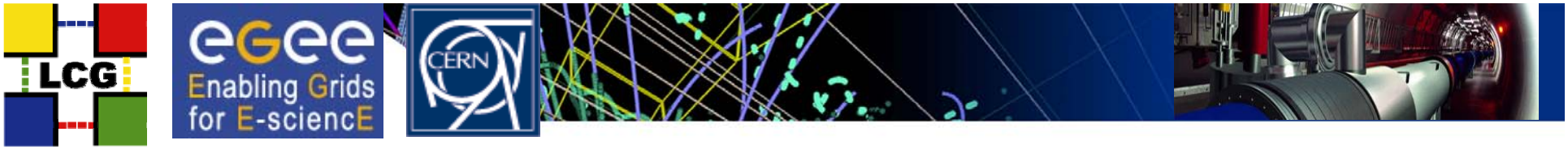
The rule of each Tier for the ALICE Experiment

- The difference between T1, T2 and T0 is just a matter of QoS (Quality of Service) for ALICE



Job Submission Structure





5th Part

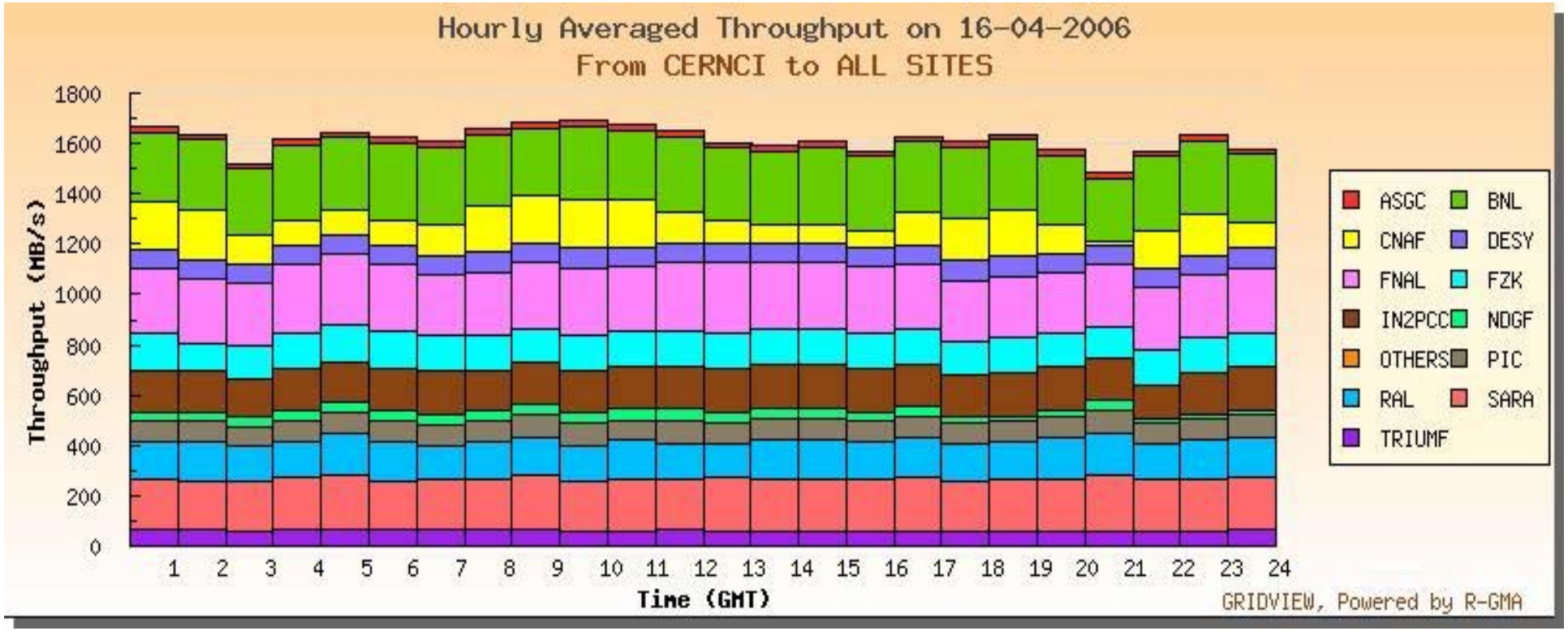
- Current Status and Applications beyond HEP

Grid Status

- WLCG Services and sites
  - The LCG Service has been validated over the past 2 years via a series of dedicated "Service Challenges", designed to test the readiness of the service infrastructure
  - These are complementary to tests by the experiments of the offline Computing Models – the Service Challenges have progressively ramped up the level of service in preparation for ever more detailed tests by the experiments
- ➢ The target: full production services by end September 2006!
- Some additional functionality is still to be added, resource levels will continue to ramp-up in 2007 and beyond
- Resource requirements are strongly coupled to total volume of data acquired to date

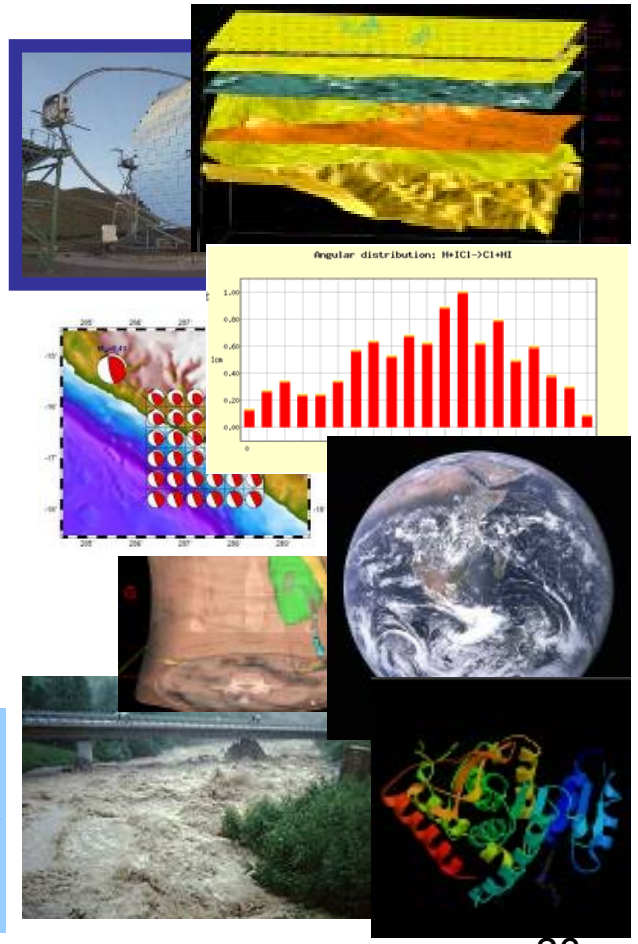
Service Challenge

- Significant focus on Data Management, including data export



Beyond HEP

- Due to the computational power of the EGEE new communities are requiring services for different research fields
- Normally these communities do not need the complex structure that required by the HEP communities
  - In many cases, their productions are shorter and well defined in the year
  - The amount of CPU required is much lower and also the Storage capabilities



20 applications from 7 domains

High Energy Physic, Biomedicine, Earth Sciences, Computational Chemistry
Astronomy, Geo-physics and financial simulation

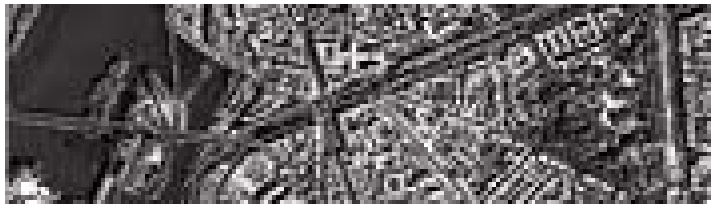
Beyond HEP: UNOSAT

UNOSAT: United Nations Initiative

- Objectives:
  - Provide the humanitarian community with access to satellite imagery and Geographic Information System services
  - Ensure cost-effective and timely products

- Core Services:
  - Humanitarian Mapping
  - Images Processing

VEGETATION – 1 Km

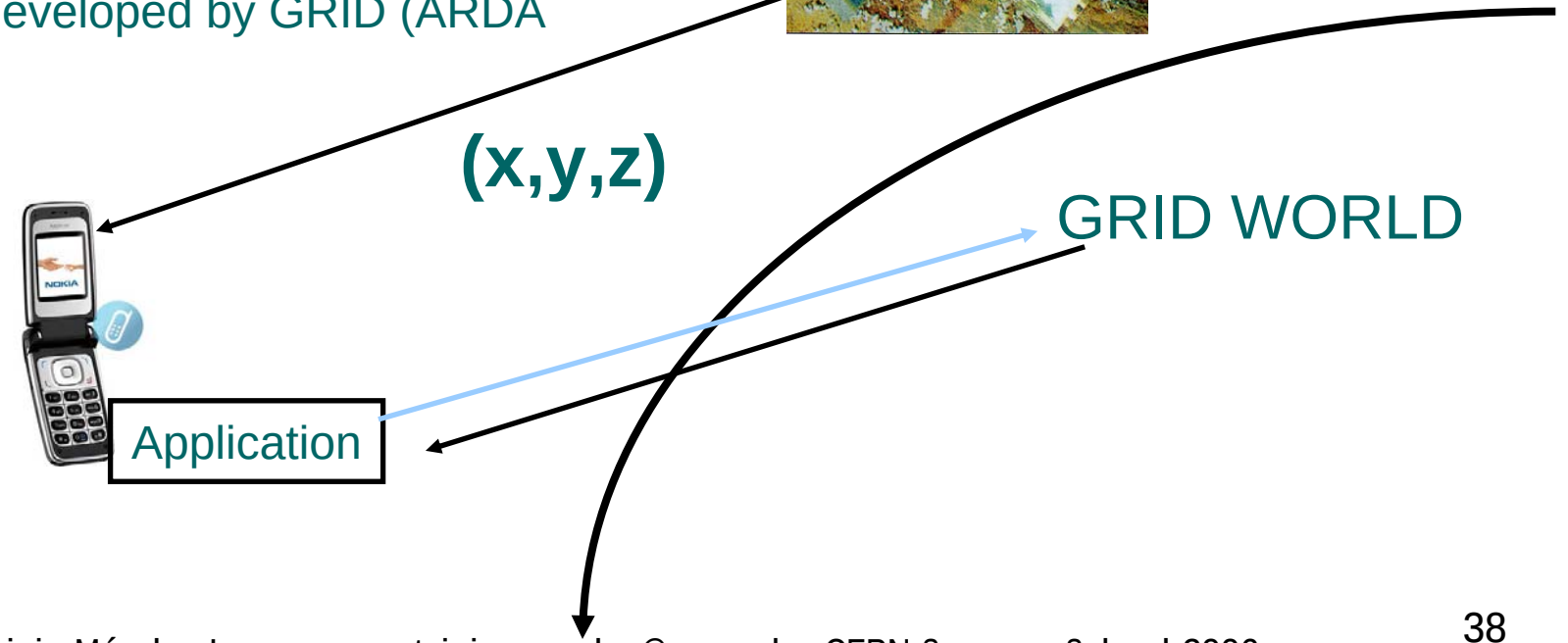


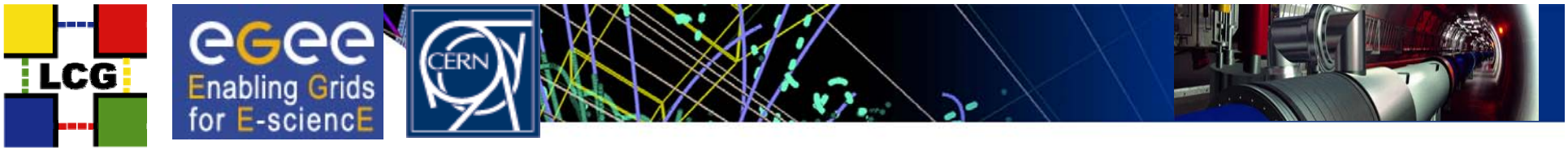
IKONOS – 1m

UNOSAT+GRID Project

- Collaboration between UNOSAT, GRID teams and EnginFrame

Fundamental actor: an application able to map metadata (coordinates) to physical location of the files: AMGA application developed by GRID (ARDA GROUP)





Production with the ITU

- May 2006: The **International Telecommunication Union** organized a world conference to establish establish a new frequency plan for the introduction of digital broadcasting in Europe, Africa, Arab States and former Russian Federation States
- The software developed by the European Broadcasting Union (EBU), performs compatibility analysis between digital requirements and existing analogue broadcasting stations
- The peculiarity is that the CPU of each job is not easily predictable, since it depends on the details of the input data set although the total one could be calculated
- The duration of the jobs variates from few seconds until several hours
- The GRID was used during the whole production to executed all the required jobs (several thousands per weekend)

Final Consideration

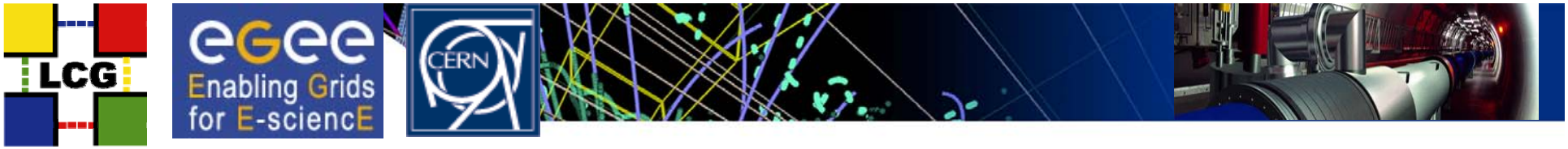
- Users have to be aware that GRID is not the magic bullet
  - Be sure what you want to do with the system
  - If you want to run few short jobs, maybe GRID is not your solution
- You will have to face situations as:
  - CPU at CERN, data in Taipei, RB in Italy….

```
[lxplus094] ~ > ping adc0018.cern.ch
PING adc0018.cern.ch (137.138.225.48) from 137.138.4.103 : 56(84) bytes of data.
--- adc0018.cern.ch ping statistics ---
11 packets transmitted, 11 received, 0% loss, time 10099ms
rtt min/avg/max/mdev = 0.204/0.405/1.332/0.332 ms
```

O,4ms

```
[lxplus094] ~ > ping lcg00105.grid.sinica.edu.tw
PING lcg00105.grid.sinica.edu.tw (140.109.98.135) from 137.138.4.103 : 56(84) bytes of data.
--- lcg00105.grid.sinica.edu.tw ping statistics ---
10 packets transmitted, 10 received, 0% loss, time 9086ms
rtt min/avg/max/mdev = 301.246/301.342/301.837/0.714 ms
```

301ms



Summary

- WLCG has been born to cover the high computer and storage demand of the experiments from 2007
- We count today more than 200 sites in 34 countries
- The whole project is the results of a big collaboration between experiments, developers and sites
- It is a reality, GRID is being used in production
- Covering a wide range of research fields
- If we were able to share information with the WEB, GRID will allow us to share computational and storage power all over the world