eGee

Enabling Grids for E-sciencE

gLite/EGEE in Practice

Alex Villazon (DPS, Innsbruck)

Markus Baumgartner (GUP, Linz)

ISPDC 2007

5-8 July 2007

Hagenberg, Austria

www.eu-egee.org

Information Society
and Media

EGEE and gLite are registered trademarks

Enabling Grids for E-sciencE

- **Theoretical part**

  - Basic Grid services

  - EGEE II project

  - gLite middleware

    - Overview and architecture

- **Practical part**

  - Live exercises with gLite testbed

Enabling Grids for E-sciencE
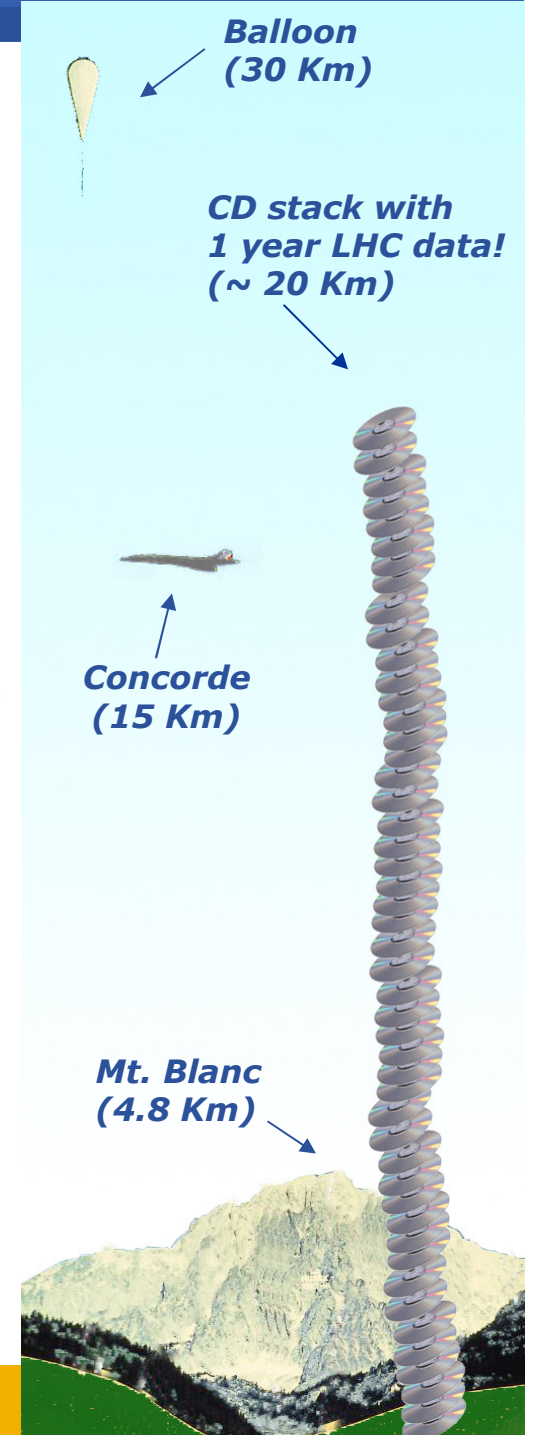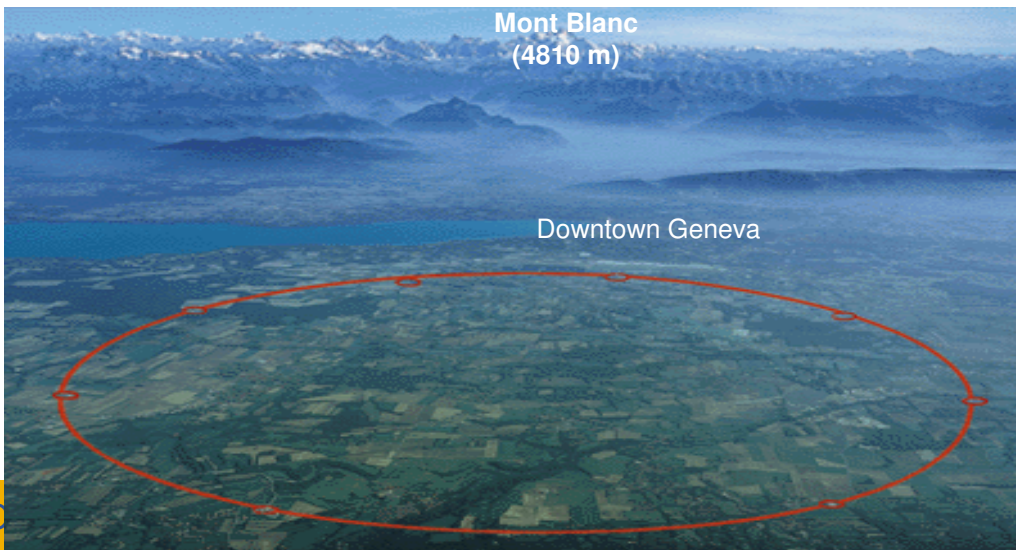
• **Why the Grid?**

- Science is becoming increasingly digital and needs to deal with increasing amounts of data
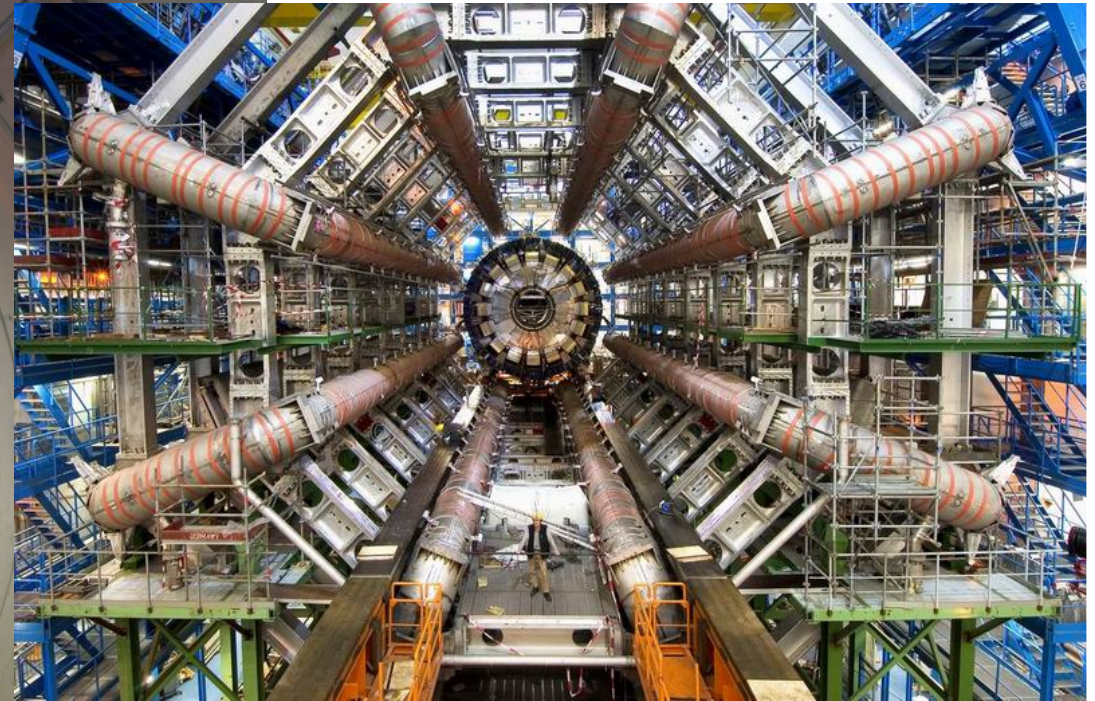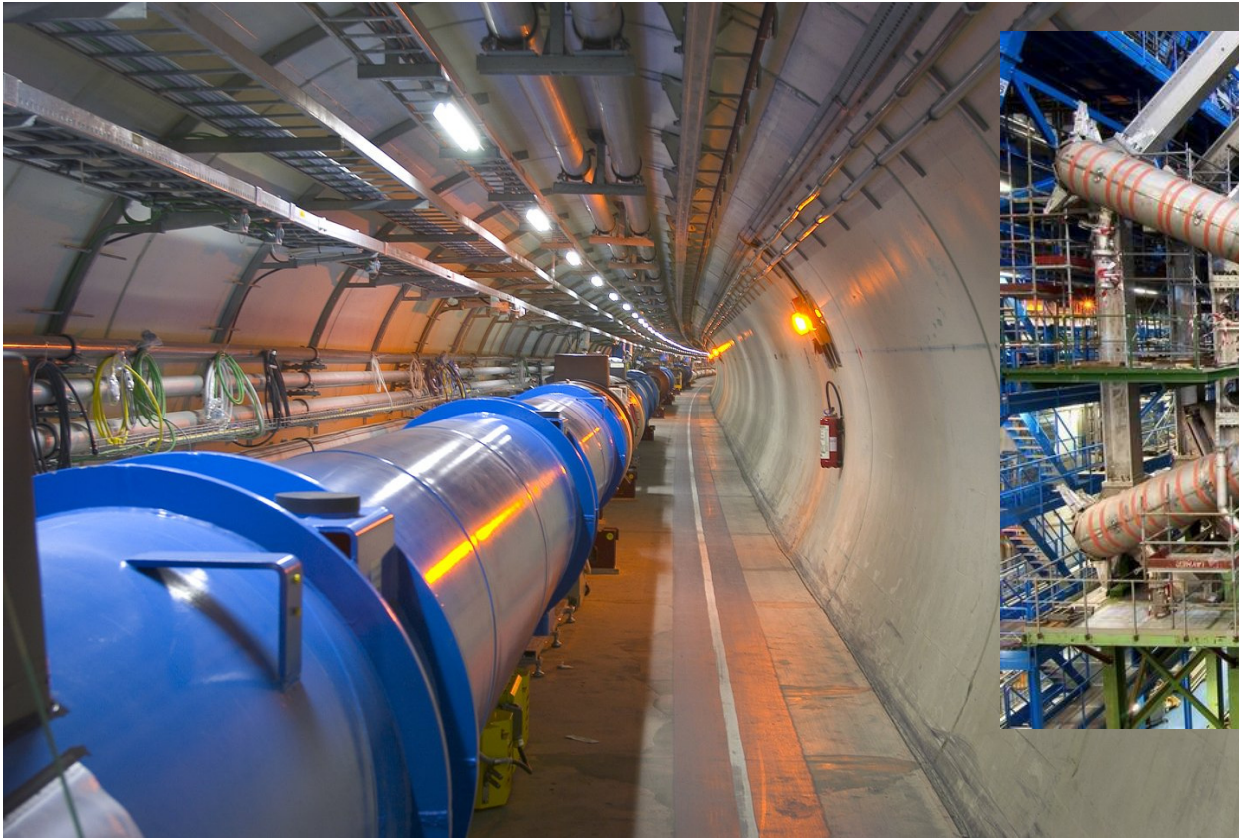 - Particle Physics and other disciplines
    - *Large amount of data produced*
    - *Large worldwide organized collaborations*
  - e.g. Large Hadron Collider (LHC) at CERN (Geneva)
    - *40 million collisions per second*
    - *~10 petabytes/year (~10 Million GBytes)*

Balloon (30 Km)

CD stack with 1 year LHC data! (~ 20 Km)

Concorde (15 Km)

Mt. Blanc (4.8 Km)

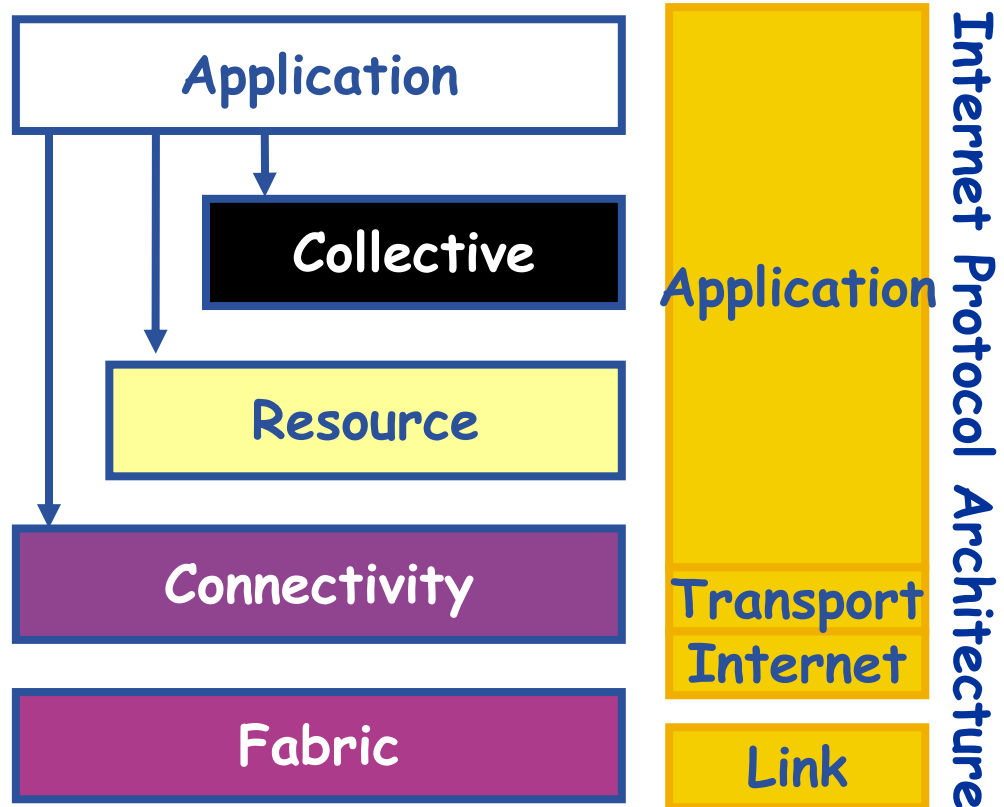**Mont Blanc (4810 m)**

Downtown Geneva

- **The biggest scientific instrument in the world starts running 2007**

… securely share distributed resources **(computation, storage, etc) so that users can** collaborate **within** Virtual Organisations (VO)

eGee
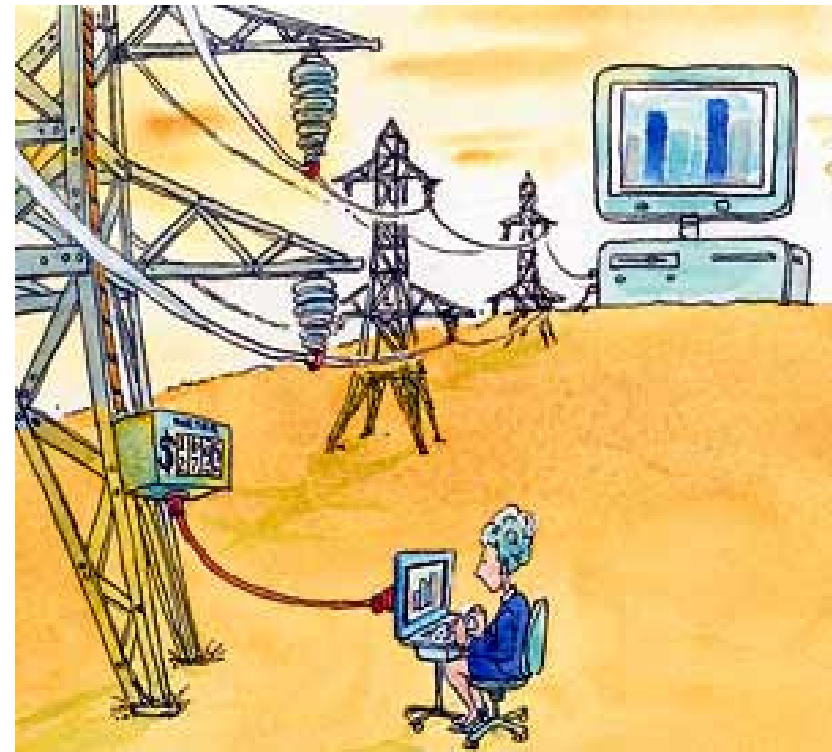
- **Application layer**
  - Grid programs
- **Collective layer**
  - Resource Co-allocation
  - Data Replica Management
- **Resource layer**
  - Resource Management
  - Information Services
  - Data Access
- **Connectivity layer**
  - Grid Security Infrastructure
  - High-performance data transfer protocols
- **Fabric layer**
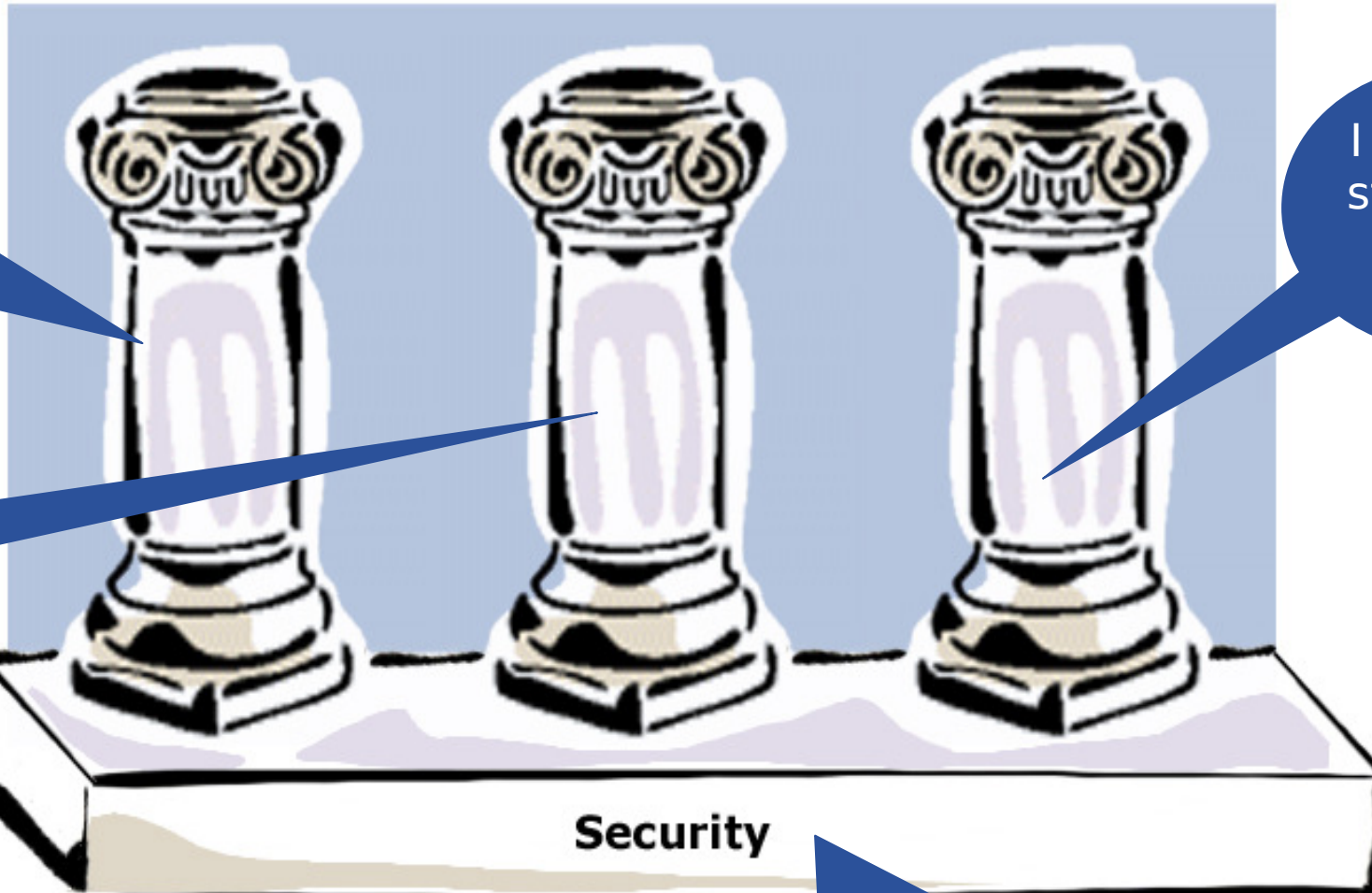 - the hardware: computers (parallel, clusters..), data storage servers

Application

Collective

Resource

Connectivity

Fabric

Internet Protocol Architecture

Application

Transport
Internet

Link

- **Defined by the Globus (http://globus.org) (Globus Toolkit)**

Enabling Grids for E-sciencE

- **Everything (or anything) is a resource**
  - Physical or logical (single computer, cluster, parallel, data storage, an application...)
  - Defined in terms of **interfaces**, not devices
- **Each site must be autonomous (local system administration policy)**
- **Grid Resource Allocation Manager (GRAM)**
  - Defines resource layer protocols and APIs that enable clients to **securely instantiate a Grid computational task** (i.e. a job)
  - Secure remote job submissions
 - Relies on local resource management interfaces

```
              ┌──────────┐
              │   GRAM   │
              └──────────┘
   ┌──────────┬─────┴─────┬──────────┐
   ▼          ▼           ▼          ▼
┌──────┐  ┌──────┐    ┌──────┐   ┌──────┐
│  LL  │  │ LSF  │    │ PBS  │   │ SGE  │
└──────┘  └──────┘    └──────┘   └──────┘
```
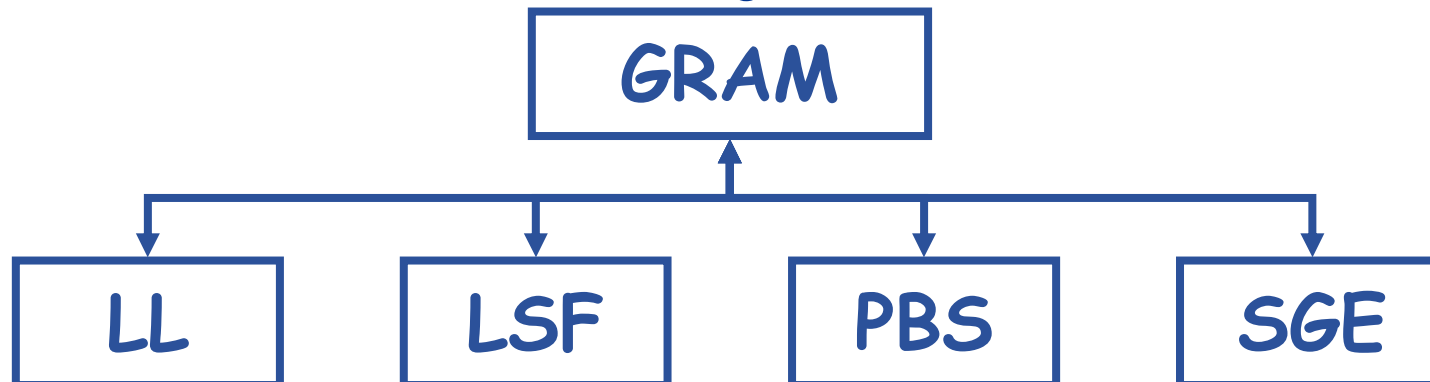
- **Job Management Services related to job management/execution**
  - Computing Element
    - job management (submission, control, …)
    - information about characteristics and status
    - Actual execution is done in a Worker Node (WN)
  - Workload Management
    - core component (see next slides)
  - Job Provenance
    - keeps track of job definition, execution conditions, environment
    - important points of the job life cycle
      - *debugging, post-mortem analysis, comparision of job execution*
  - Package Manager
    - extension of a traditional package management system to a grid
      - *automates the process of installing, upgrading, configuring and removing software packages from a shared area on a grid site*
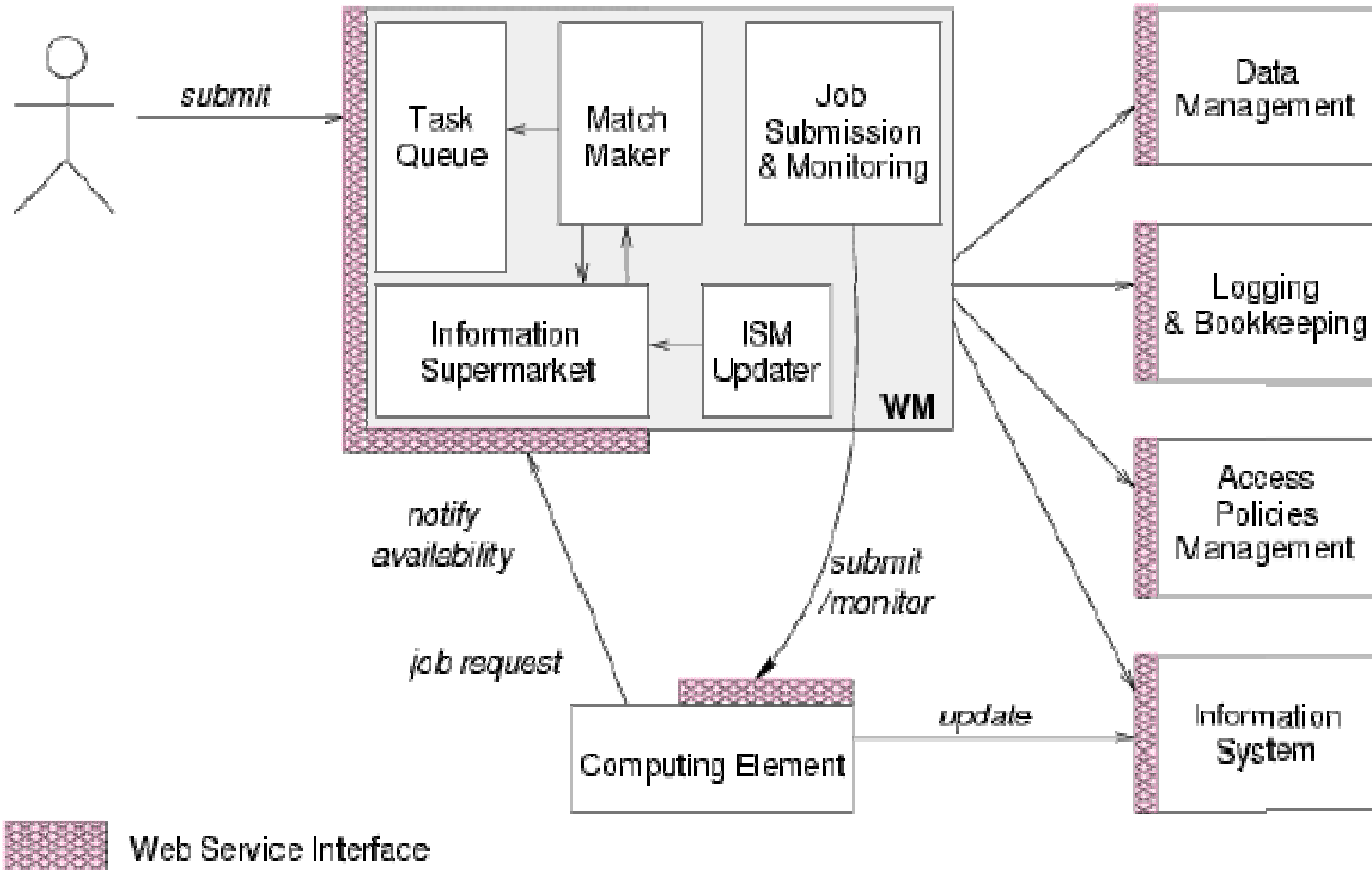
- **Maintains information about hardware, software, services and people participating in a Virtual Organization**
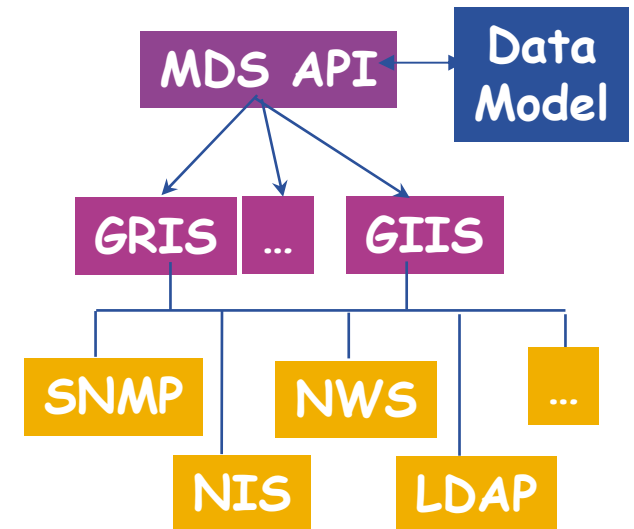 - Should scale with the Grid´s growth

"Find a computer with at least 2 free CPUs and with 10GB of free disk space..."

- **Globus MDS (Metacomputing Directory Service)**
  - Hierarchical, push based (pull based)
   ➔ showed limitations

- **Berkely Database Information Index (BDII)**
  - A Monitoring and Discovery Service (MDS) **evolution**
  - Based on LDAP (Lightweight Directory Access Protocol)
  - Central system
    - Queries servers/providers about status
    - Stores the retrieved information in a database
    - Provides the information following the GLUE Schema
- **Commands**
  - ➢ lcg-infosites –vo <your_vo> all I ce I se I lfc I lfcLocal I –is <your_bdii>

```
[gliteui] /home/martin > lcg-infosites --vo dpsgltb all –is glitece.dps.uibk.ac.at
#CPU    Free    Total Jobs    Running Waiting ComputingElement
----------------------------------------------------------
   2    2    0         0     0   glitece.dps.uibk.ac.at:2119/blah-pbs-dpsgltb
Avail Space(Kb) Used Space(Kb)  Type    SEs
----------------------------------------------------------
3172384     4664832      n.a    gliteio.dps.uibk.ac.at
```

- **Relational Grid Monitoring Architecture (R-GMA)**
  - Developed as part of the EuropeanDataGrid Project (EDG)
  - Now as part of the EGEE project
  - Based on the Grid Monitoring Architecture (GMA)

- **Uses a relational data model**
  - There is no central repository, only a "Virtual Database"
  - Schema is a list of table definitions
    - Additional tables/schema can be defined
  - Registry is a list of data producers with all its details
  - Producers publish data
    - From sites and applications
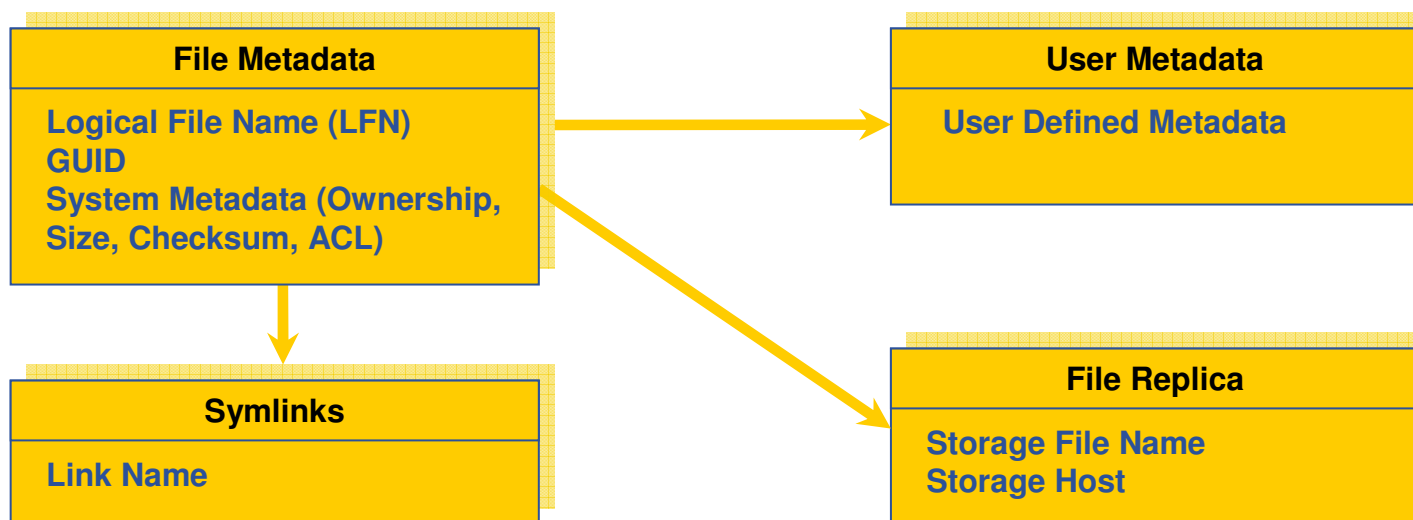 - Consumers read published data

| Prod | ... | Prod |

Schema   Virtual table   Registry

| Cons | ... | Cons |
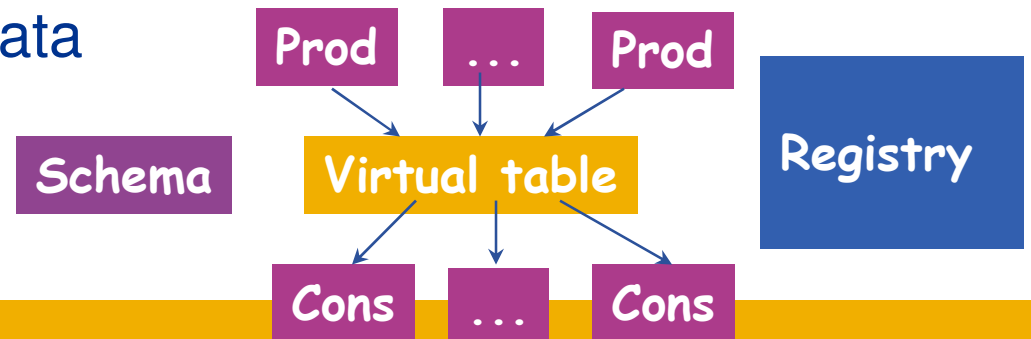
- **Data access and transfer**
  - Simple, automatic multi-protocol file transfer tools: **Integrated with Resource Management service**
    - Move data from/to local machine to remote machine, where the job is executed (staging – stageout)
    - Redirect stdin to a remote location
    - Redirect stdout and stderr to the local computer
    - Pull executable from a remote location
  - To have a secure, high-performance, reliable file transfer over modern WANs: **GridFTP**

- **User and programs produce and require data**
  - Resource Broker can send data from/to jobs
    - Input/Output Sandboxes are limited to 10 MB
   - Data has to be copied from/to local filesystems to the Grid (UI, WN)

- **Solution**
  - Storing data in Grid datasets
    - Located in Storage Elementes (SE)
    - Several replicas of one file in different sites
    - Accessible by Grid users and applications from "everywhere"
    - Locatable by the WMS (data requirements in JDL)

- **LCG File Catalog**
  - Unique Identifier (GUID)
  - One single catalog with LFN-> GUID -> SFN mapping
  - All entities are treated/replicated like files in a UNIX filesystem
  - Hierarchical namespace
  - System attributes stored as metadata on the GUID (1 field of user metadata
  - Transactions, timeoutes, retries
  - Relational database backend (Oracle and MySql)

File Metadata
Logical File Name (LFN) **GUID** **System Metadata (Ownership,** **Size, Checksum, ACL)**

User Metadata
User Defined Metadata

Symlinks
Link Name

File Replica
Storage File Name **Storage Host**

- **Catalog**
  - File and Replica Catalog
  - File Authorization Service
  - Metadata catalog
  - Distribution of catalogs, conflicts resolution
- **Storage Elements (SE)**
  - SRM (Storage Resource Manager) interface
  - Transfer protocols (**gsiftp**, rfio, …)

Catalog

SE SE SE SE SE SE
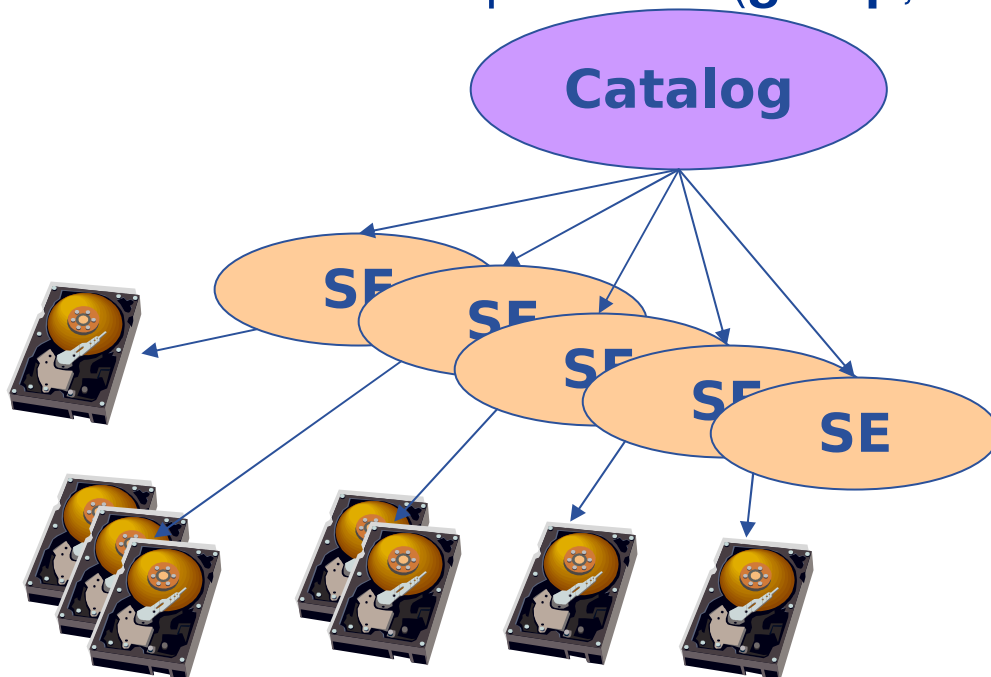
Logical File Name
LFN : /grid/gilda/basel/file.txt

Storage Resource Manager
srm://trigrid-
ce01.unime.it/dpm/unime.it/home/gilda/generated/
2006-09-20/filef026441a-5834-431f-b28d-06cb7e4c784f

Physical Filename
/home/gilda/generated/2006-09-20/filef026441a-5834-
431f-b28d-
06cb7e4c784f

- Logical File Name (LFN)
  - An alias created by user to refer some data item
    "lfn:/grid/dpsgltb/20070609/test/example.txt"

- Globally Unique Identifier (GUID)
  - A non-human-readable unique identifier
    "guid:f813d4ac-7dec-32f0-00aa09bfe6ec
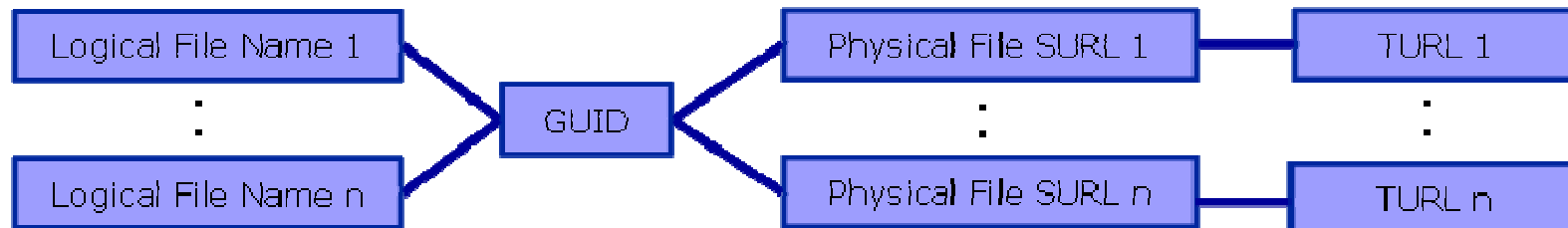
- Site URL (SURL)
  - Location of data on a storage system
    "srm://gliteio.dps.uibk.ac.at/files/dpsgltb/output7_3" (SRM)
    "sfn://gliteio.dps.uibk.ac.at/storage/dpsgltb/file10.dat" (Classic SE)

- Transport URL (TURL)
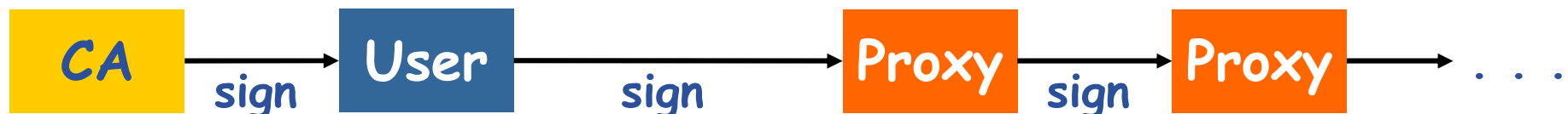 - Temporary locator of a replica + access protocol
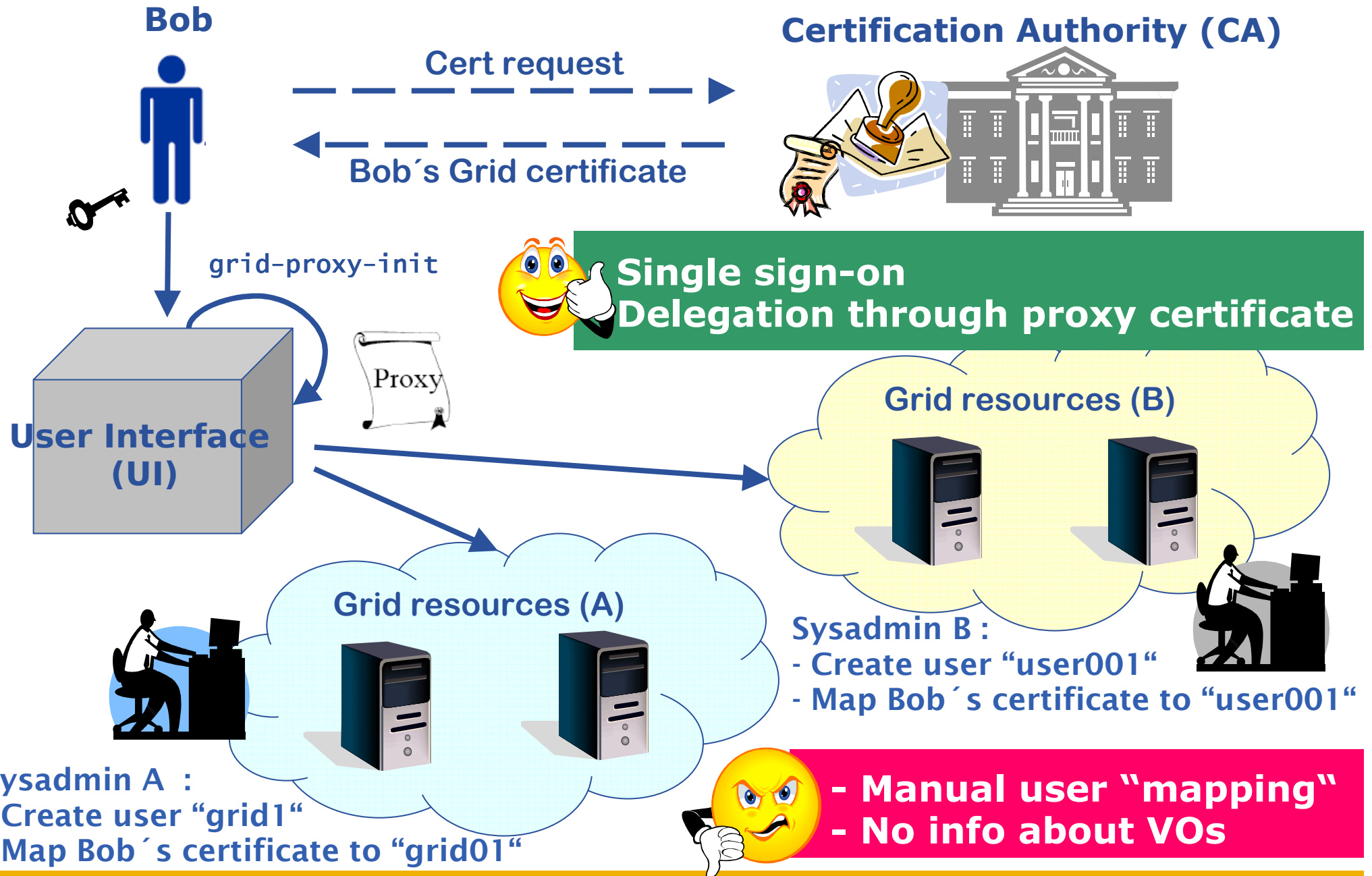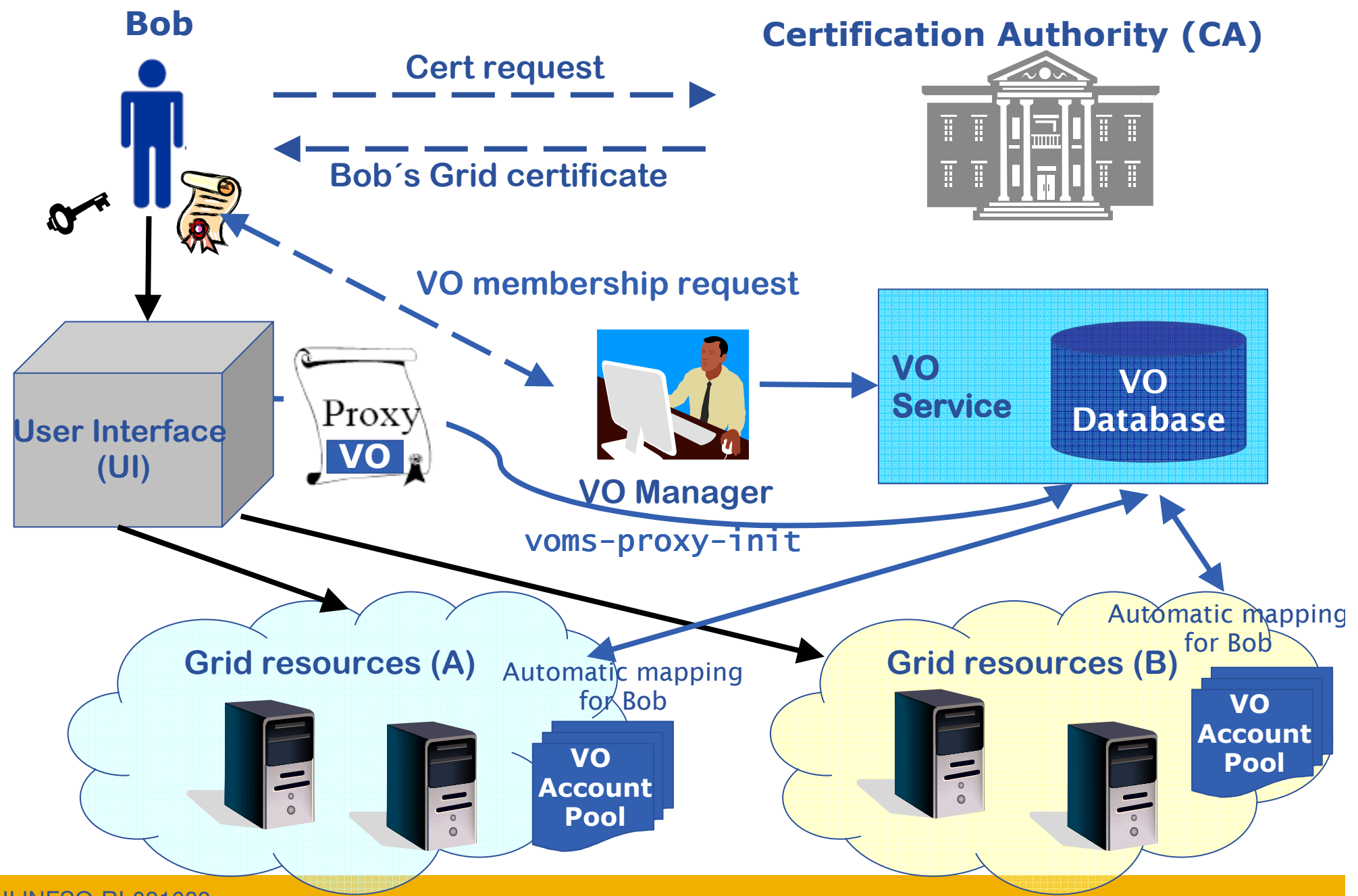   "rfio://gliteio.dps.uibk.ac.at//storage/dpsgltb/file10.dat"

lfc-chmod	Change access mode of the LFC file/directory
lfc-chown	Change owner and group of the LFC file-directory
lfc-delcomment	Delete the comment associated with the file/directory
lfc-getacl	Get file/directory access control lists
lfc-ln	Make a symbolic link to a file/directory
lfc-ls	List file/directory entries in a directory
lfc-mkdir	Create a directory
lfc-rename	Rename a file/directory
lfc-rm	Remove a file/directory
lfc-setacl	Set file/directory access control lists
lfc-setcomment	Add/replace a comment

[rabmar95@glite-tutor tmp]$ **lcg-cr -v --vo gilda file:/tmp/dummy.tar.gz -d trigriden01.unime.it -l lfn:/grid/gilda/tmp/dummy.tar.gz**
Using grid catalog type: lfc
Using grid catalog : lfc-gilda.ct.infn.it
Using LFN : /grid/gilda/tmp/dummy.tar.gz
Using SURL : sfn://trigriden01.unime.it/flatfiles/SE00/gilda/generated/2007-06-11/file20bf7537-d6d6-47a6-91bc-6f47314568b4
Source URL: file:/tmp/dummy.tar.gz
File size: 154
VO name: gilda
Destination specified: trigriden01.unime.it
Destination URL for copy: gsiftp://trigriden01.unime.it/flatfiles/SE00/gilda/generated/2007-06-11/file20bf7537-d6d6-47a6-91bc-6f47314568b4
streams: 1
set timeout to 0 seconds
Alias registered in Catalog: lfn:/grid/gilda/tmp/dummy.tar.gz
        154 bytes   0.44 KB/sec avg   0.44 KB/sec inst
Transfer took 1040 ms
Destination URL registered in Catalog: sfn://trigriden01.unime.it/flatfiles/SE00/gilda/generated/2007-06-11/file20bf7537-d6d6-47a6-91bc-6f47314568b4
guid:47145cba-1d99-46f3-9c43-fc5267add103

Enabling Grids for E-sciencE

- **Basic security:**
  - **Authentication:** Who we are on the Grid?
  - **Authorization:** Do we have access to a resource/service?
  - **Protection:** Data integrity and confidentiality
- **but, there are thousands of resources over different administration domains...:**
  - **Single sign-on**, i.e. give a password once, and be able to access all resources (to which we have access)
- **Grid Security Infrastructure (GSI):**
  - **Grid credentials:** digital certificate and private key
    - Based on Public Key Infrastructure (PKI). X.509 standard
    - Certification Authority (CA) signs certificates. Trust relationship
  - **Proxy certificates:** Temporary self-signed certs, allowing single sign-on: ***Proxy delegation***

```
CA --sign--> User --sign--> Proxy --sign--> Proxy --> . . .
```

Bob

Certification Authority (CA)

Cert request

Bob´s Grid certificate

User Interface (UI)

Proxy
VO

VO membership request

VO Manager

voms-proxy-init

VO Service

VO Database

Grid resources (A)

Automatic mapping for Bob

VO Account Pool

Grid resources (B)

Automatic mapping for Bob

VO Account Pool

eGee

- **Virtual Organization Membership Service (VOMS)**

  – EGEE/gLite enhancement for VO management

  - Provides information on user's relationship with Virtual Organization (VO)
    - Membership
    - Group membership
    - Roles of user

- **Multiple VO**
  - User can register to multiple VOs and create an aggregate proxy
  - Access ressources in every registered VO

- **Backward compatibility**
  - Extra VO related information in users proxy certificate
 - Users proxy can still be used with non VOMS-aware services

- **Requires a valid certificate from a recognized CA imported on the browser**

- **VO user can**

  - Query membership details
  - Register himself in the VO
    - Needs a valid certificate
  - Track his requests

- **VO manager can**

  - Handle requests from users
  - Administer the VO

- **Everybody can**

  - Get information about the VO
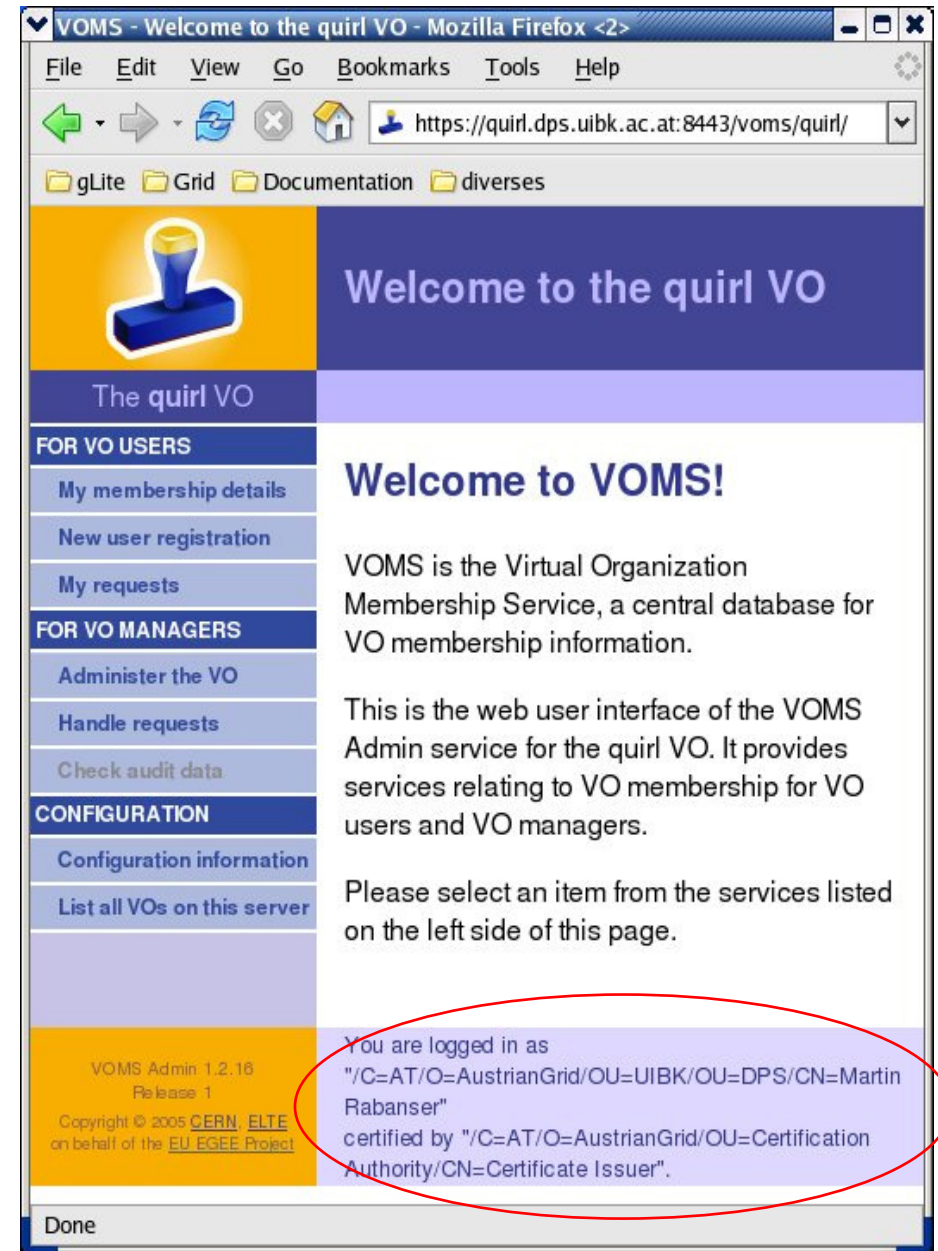
- **EGEE = Enabling Grids for E-sciencE**

  - Biggest Grid worldwide
  - 90 Million EURs project (2 years)
  - over 90 leading institutions in more than 30 countries, federated in regional Grids
  - Currently
    - 20.000 CPUs
    - 5 Petabytes (5 Mio. GB) storage
    - ~200 Virtual Organizations (VO)

CERN
Central Europe
France
Germany and Switzerland
Ireland and UK
Italy
Networking
Northern Europe
Russia
South-East Europe
South-West Europe
USA

- **Particle Physics**
- **Bioinformatics**
- **Industry**
- **Astronomy**
- **Chemistry**
- **Earth Observation**
- **Geophysics**
- **Biodiversity**
- **Nanotechnology**
- **Climate Modeling**
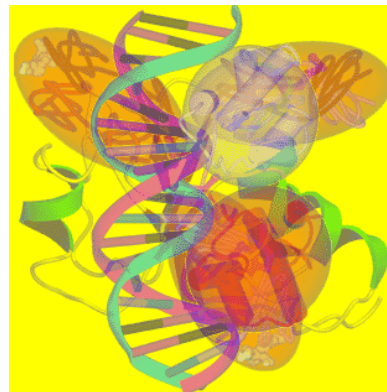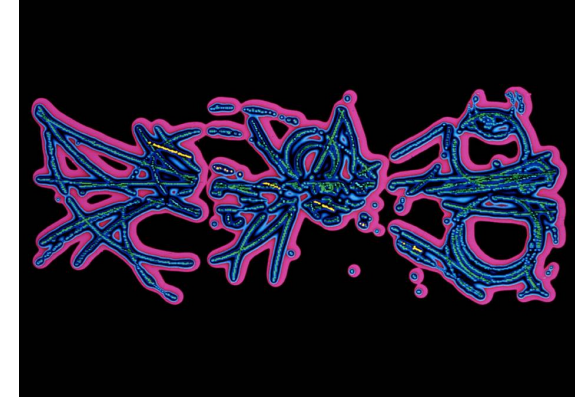
The Grid Live

Real Time Monitoring

http://gridportal.hep.ph.ic.ac.uk/rtm/

- **The Grid relies on advanced software – the middleware - which interfaces between resources and the applications**
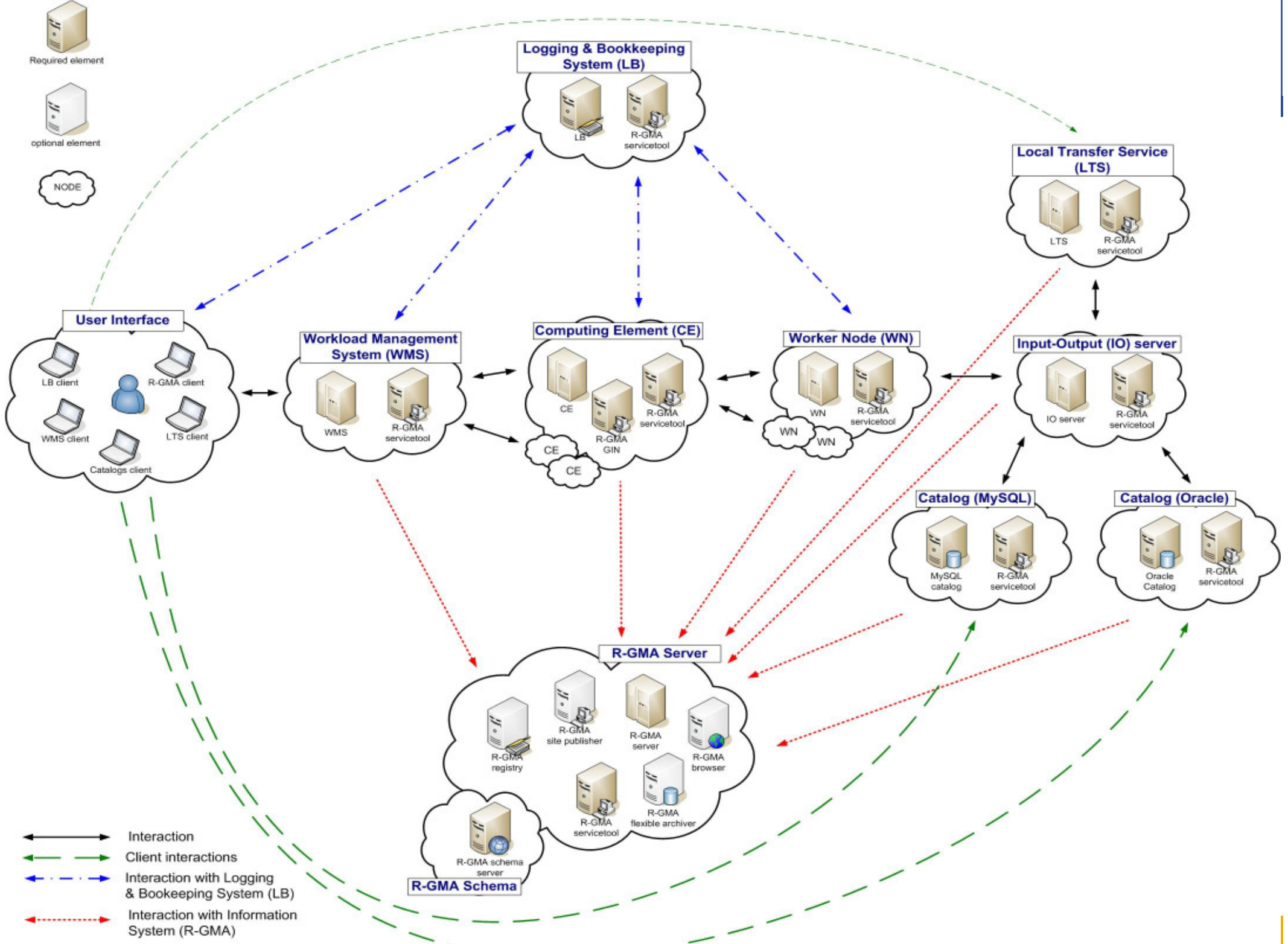
- **The GRID middleware**

  - Finds convenient places for the application to be executed
  - Optimises use of resources
  - Organises efficient access to data
  - Deals with authentication to the different sites that are used
  - Run the job & monitors progress
  - Transfers the result back to the scientist

Enabling Grids for E-sciencE
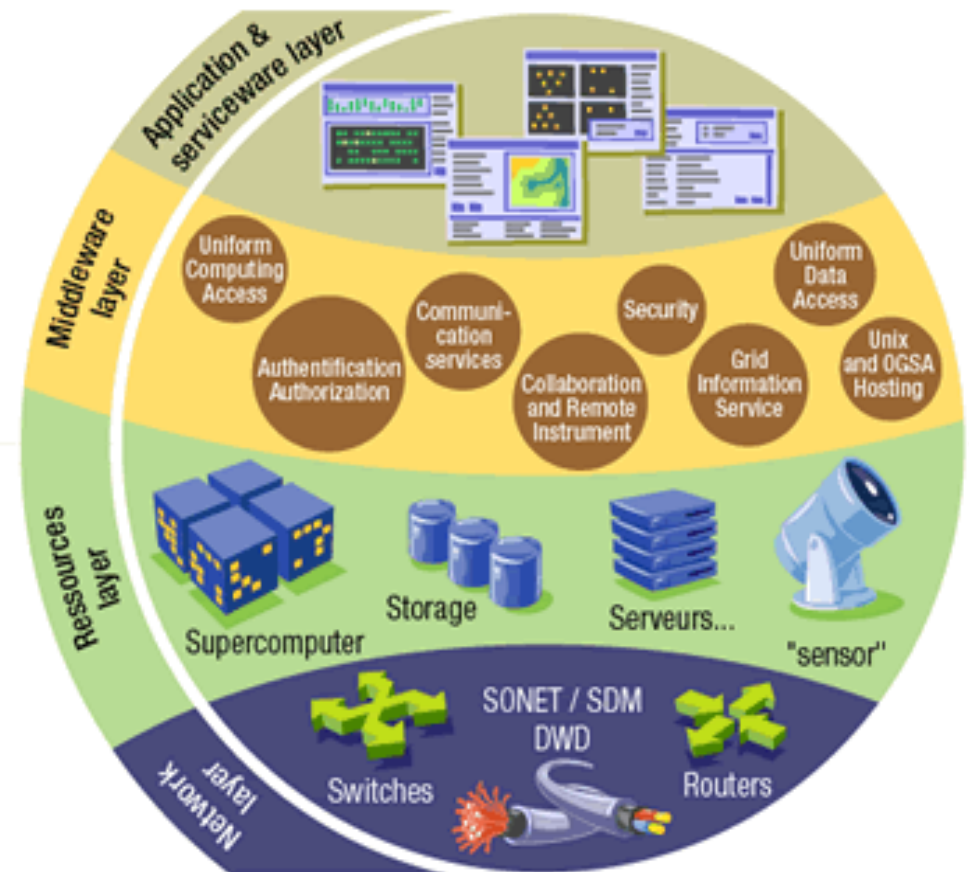
- **gLite**

  - First release 2005 (currently gLite 3.0)
  - Next generation middleware for grid computing
  - Developed from existing components (globus, condor,..)
  - Intended to replace present middleware with production quality services
  - Interoperability & Co-existence with deployed infrastructure
  - Robust: Performance & Fault tolerance
  - Open Source license

END OF FIRST PART