Introduction aux grilles de calcul et au projet EGI

David Bouvet

CCIN2P3, 18-19/10/10

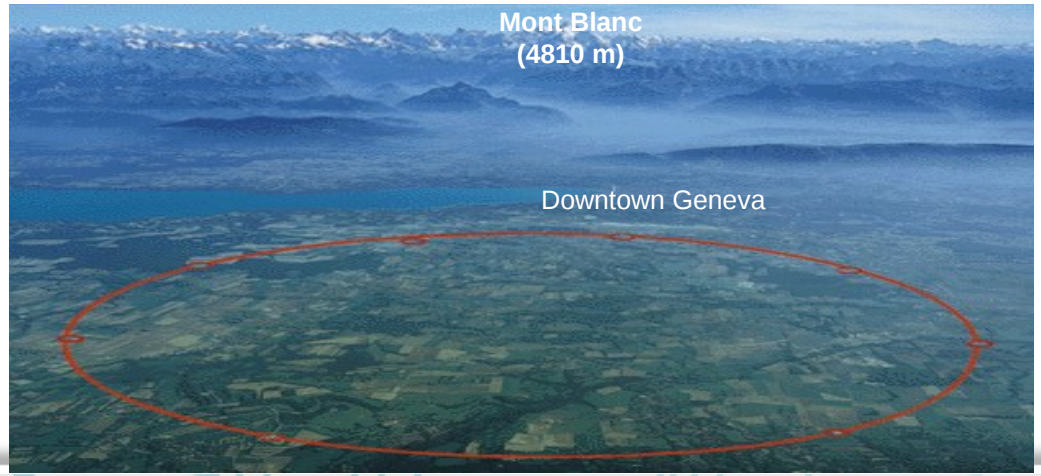
Plan

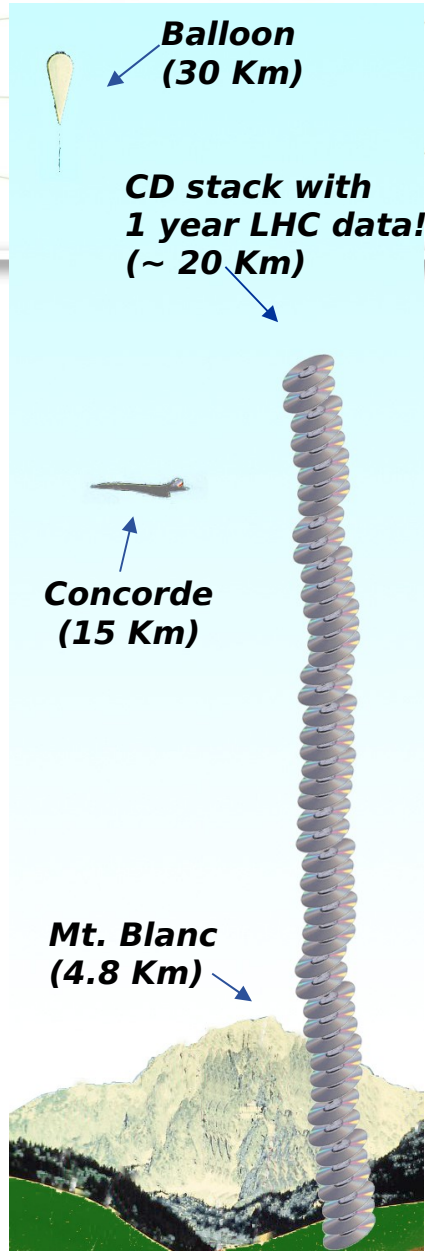
- **Introduction à la grille**
- **Le projet EGI**
- **Le middleware gLite**
  - Vue d'ensemble et architecture

Motivation

Why the grid?

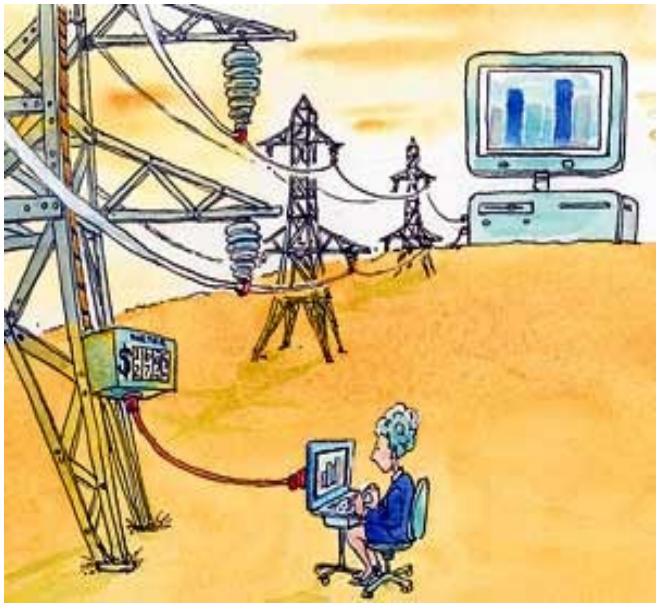
- Science is becoming increasingly digital and needs to deal with increasing amounts of data
  - Large amount of data produced
  - Large worldwide organized collaborations
    - e.g. Large Hadron Collider (LHC) at CERN (Geneva) : ~10 petabytes/year (~10 Million GBytes)





The solution: the Grid?

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



Or the GridS?

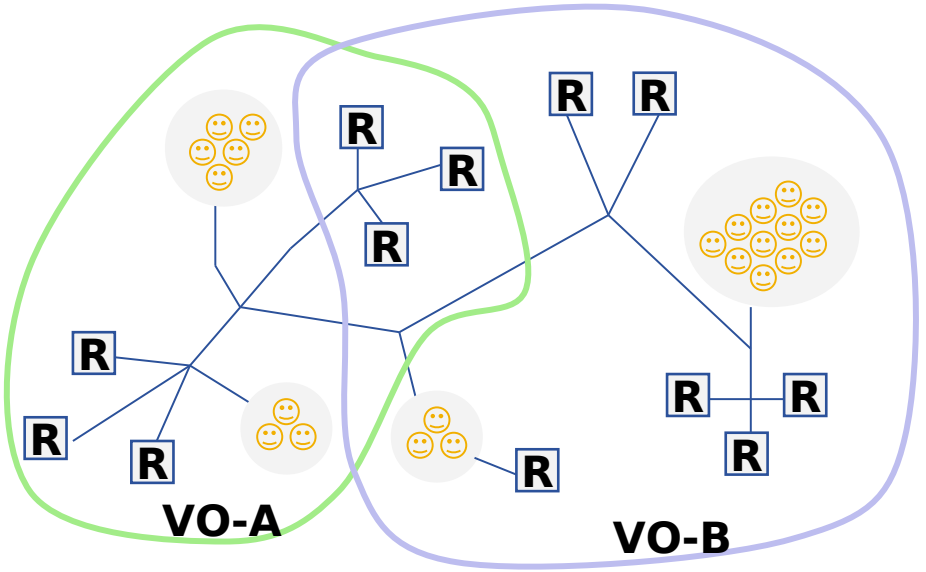
- **Various scopes: production grids/research grids**
- **More or less homogeneous grids:**
  - desktop grids
    - BOINC / @home
    - XtreemOS
  - service grids
    - EGEE / OSG / NAREGI ...
    - decrypthon
  - HPC grids
    - DEISA, TeraGrid, ...
- **eventually interoperable.**

Virtual Organisation

- **On many Grids, users are grouped in Virtual Organisations:**
  - A group of people from different institutions working on a common goal
  - Sharing distributed processing and data resources
    - Computers
    - Data files
    - Scientific instruments
    - Codes
    - ...

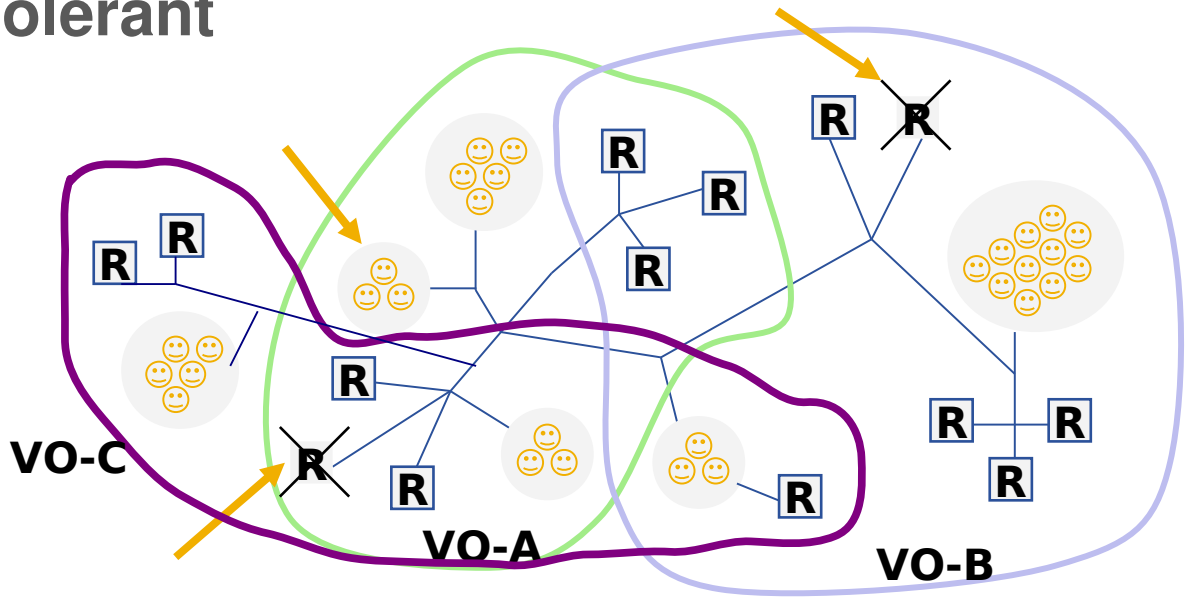
Virtual Organisation

- Distributed resources and people
- Linked by networks, cross-administrative domains
- Sharing resources, common goals



Virtual Organisation

- Distributed resources and people
- Linked by networks, cross-administrative domains
- Sharing resources, common goals
- Dynamic, fault tolerant



VOMS - What is VOMS?

- **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
  - Has an Account Database
    - Comparable to a Kerberos server
  - Serving information in a special format (VOMS credentials)
  - Administration via command line & web interface

VOMS - Features

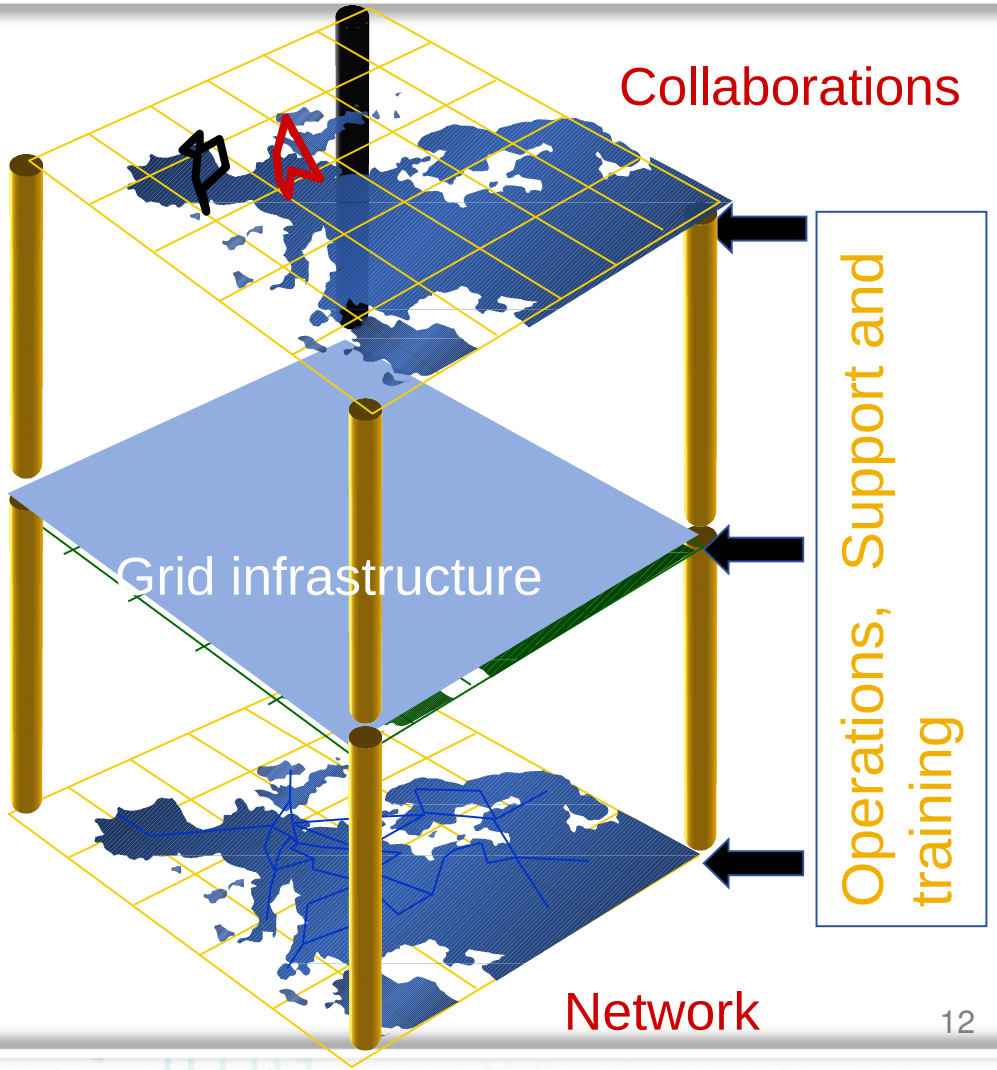
- **Single login creating a proxy at the beginning of a session**
  - Attaches VOMS certificate to user proxy certificate
- **Expiration time**
  - Authorization valid for a limited time (may differ from proxy duration)
- **Multiple VO**
  - User can register to multiple VOs
- **Backward compatibility**
  - Extra VO related information in users proxy certificate
  - Users proxy can still be used with non VOMS-aware services
- **Security**
  - Client/Server communication are secured and authenticated

EGEE

- **EGEE : Enabling Grids for E-sciencE**
  - The flagship Grid Infrastructure project of the EU
  - Funded by the European Commission
  - Primary Objectives
    - consistent, robust and secure service grid infrastructure
    - improving and maintaining the middleware
    - attracting new users from industry and science
      - ensure they receive high standard of training and support
  - Structure
    - over 250 computing centers in more than
    - 48 countries, federating in regional Grids
    - 60.000 CPUs, > 5 Petabytes storage

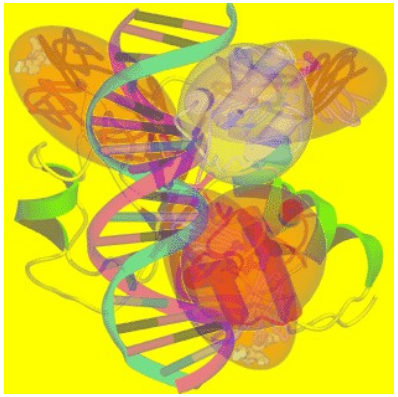
What is the EGEE project?

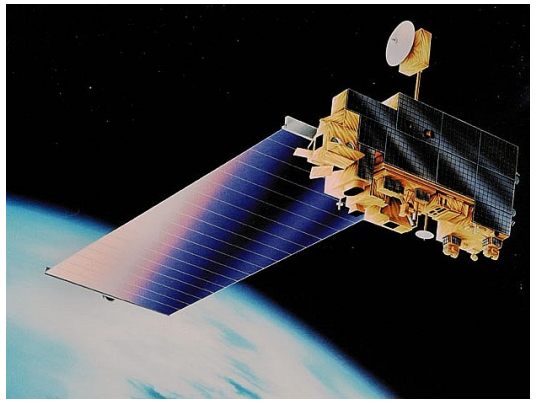
- **Build a large-scale production grid service to:**
  - Support science and technology worldwide
  - Foster international cooperation both in the creation and the use of the e-infrastructure
  - Link with and build on national, regional and international initiatives



Applications in EGEE

- Particle Physics
- Bioinformatics
- Industry
- Astronomy
- Chemistry
- Earth Observation
- Geophysics
- Biodiversity
- Nanotechnology
- Climate Modeling





Kind of applications

- Exploring big parameter space
- Deterministic / probabilistic
- “Gridifying “ legacy code / designed for Grids
- but also:
- standalone / parallel (MPI)
- short runs / long runs (checkpoints)

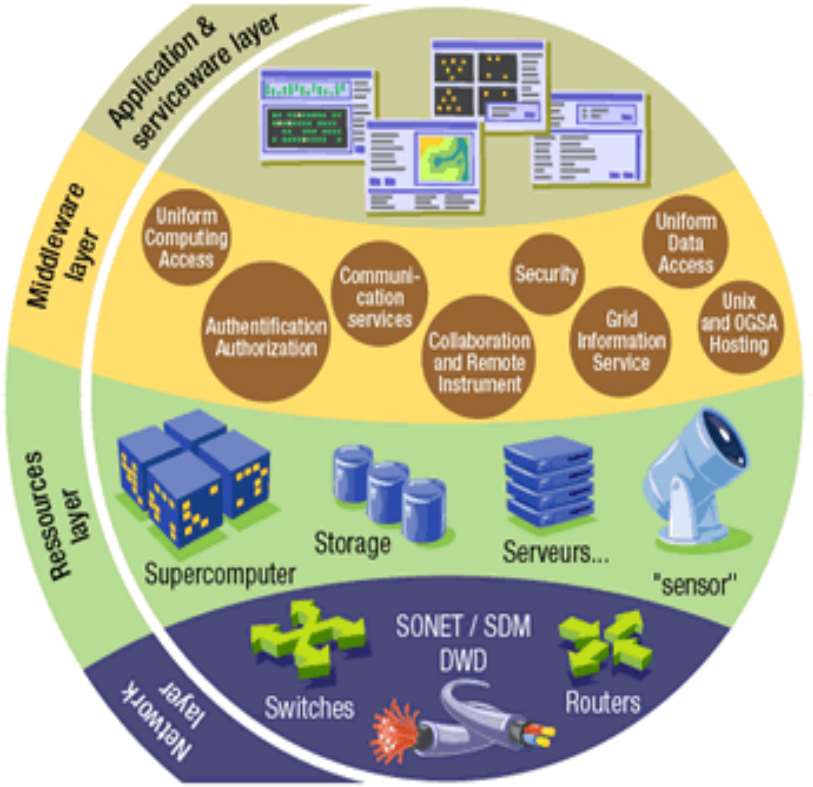
- Often complex data-flow and control-flow
- (Grid workflow applications)

EGI

- **Suite du projet EGEE, il vise à pérenniser la grille européenne.**
  - À débuter en avril 2010
- **Basé sur les NGI (National Grid Initiative)**

gLite – Grid middleware

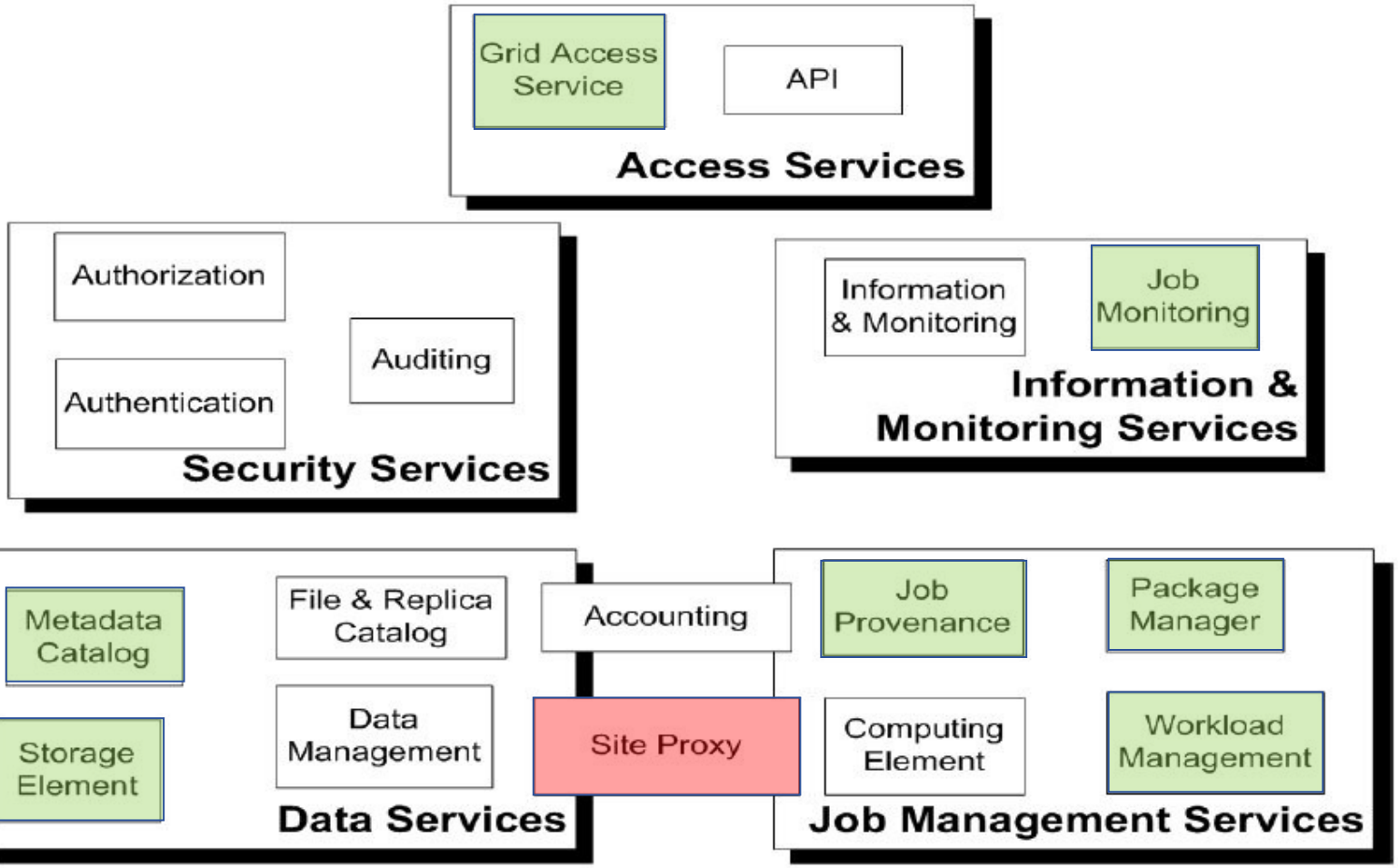
- **The Grid relies on advanced software – the middleware - which interfaces between resources and the applications**
- **The GRID middleware**
  - Finds convenient places for the application to be executed
  - Optimises use of resources
  - Organises efficient access to data
  - Deals with authentication to the different sites that are used
  - Run the job & monitors progress
  - Transfers the result back accounts all operations



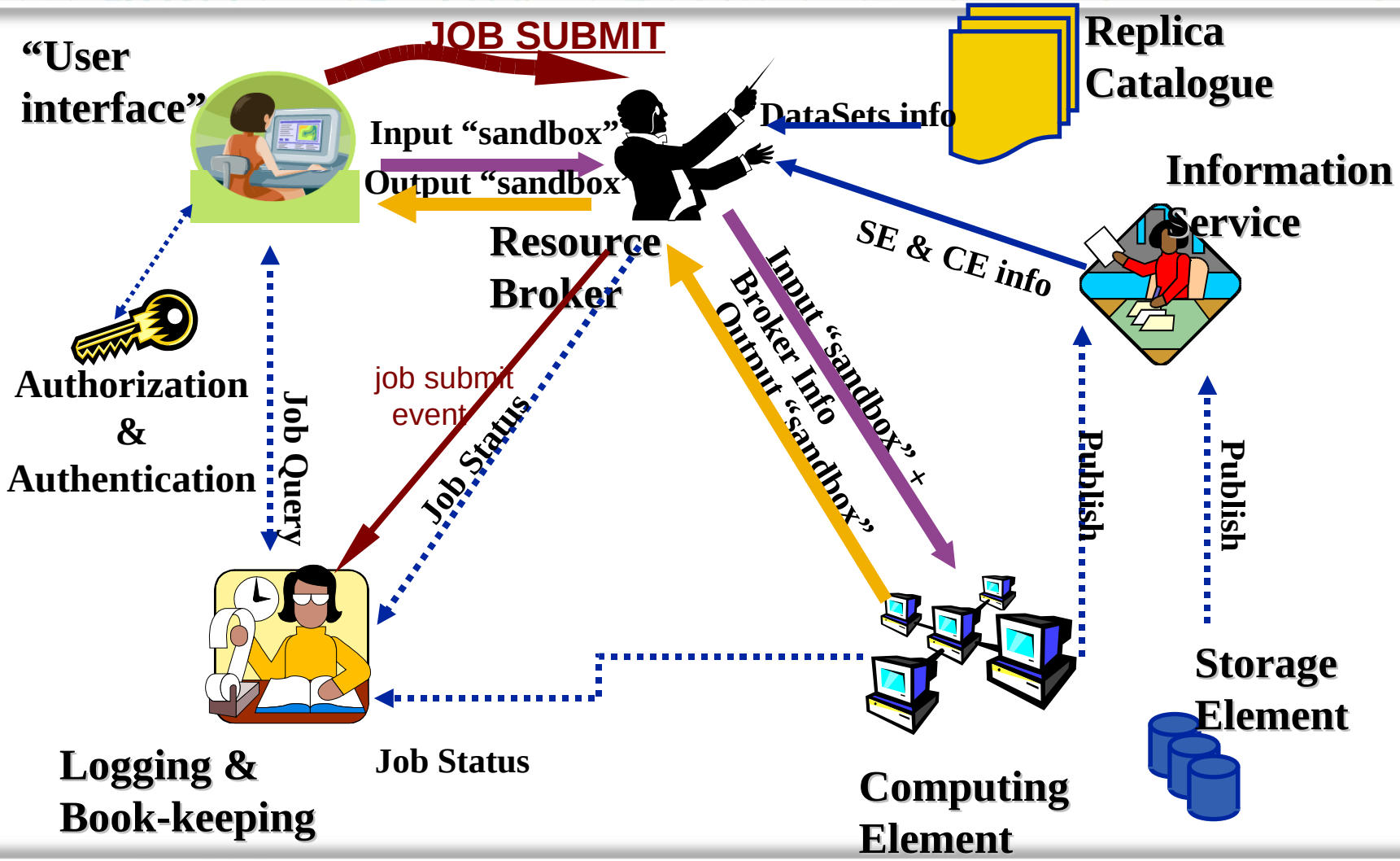
gLite – Overview

- **gLite**
  - First release 2005 (currently gLite 3.2)
  - Next generation middleware for grid computing
  - Intended to replace present middleware with production quality services
  - Developed from existing components (globus, condor,..)
  - Interoperability & Co-existence with deployed infrastructure
  - Robust: Performance & Fault tolerance
  - Open Source license
  - Platform: Currently only Scientific Linux supported

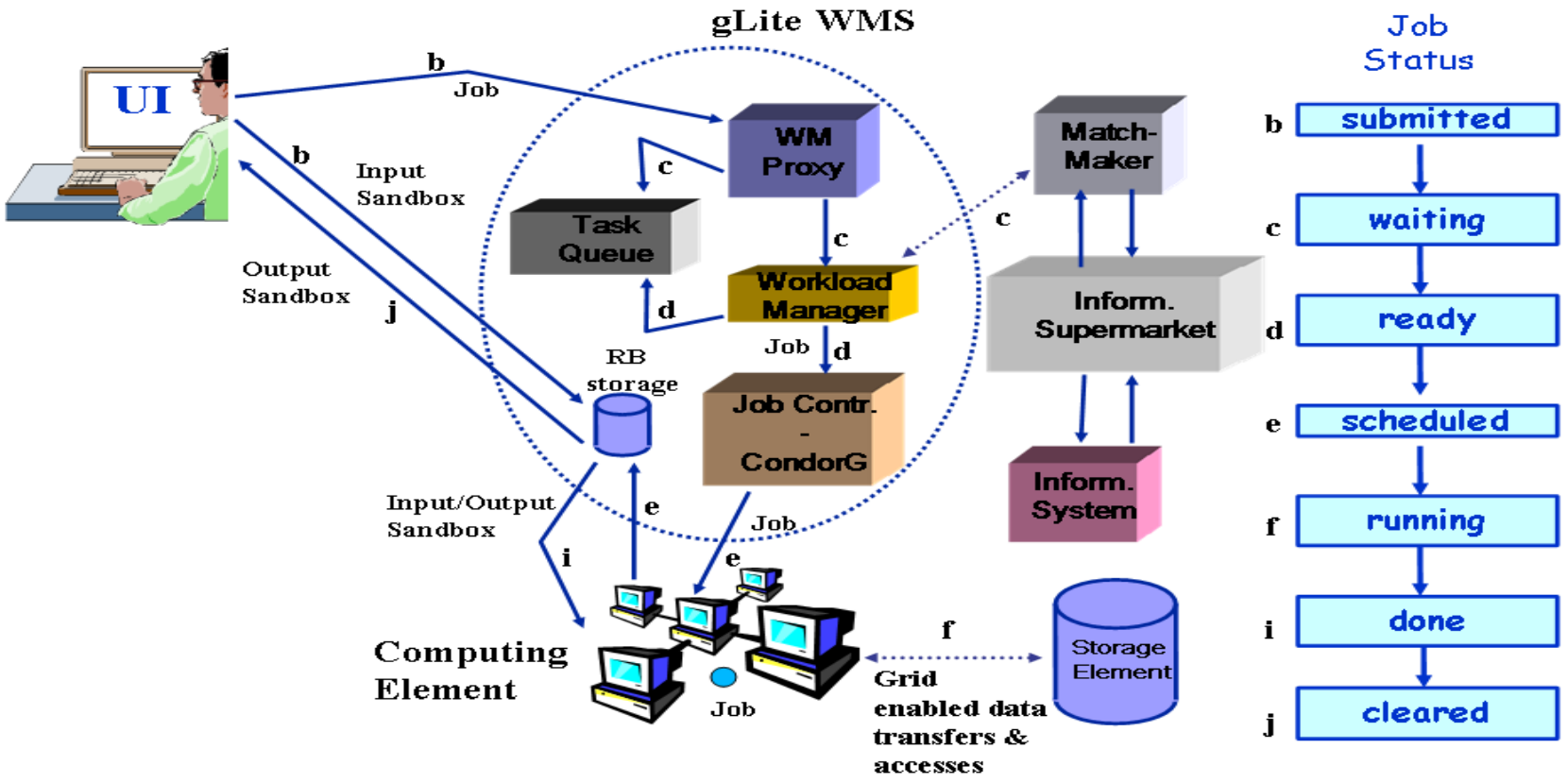
gLite – Services



gLite – Services



gLite WMS - Job flow



Other middleware services

- **File Transfer Service**
- **Metadata management (interface to databases management systems)**
  - AMGA
  - GRELC / GDSE / OGSA-DAI
- **Alternate WMS**
  - gridway

Links

- **gLite user documentation**
  - https://edms.cern.ch/document/722398/
- **Troubleshooting**
  - https://gus.fzk.de
- **Job monitoring**
  - http://gridview.cern.ch
- **Grid monitoring**
  - http://gridportal.hep.ph.ic.ac.uk/rtm/
  - http://gstat-prod.cern.ch/gstat