Advanced School in High Performance and GRID Computing – Concepts and Applications, ICTP, Trieste, Italy

Introduction to GRID Computing

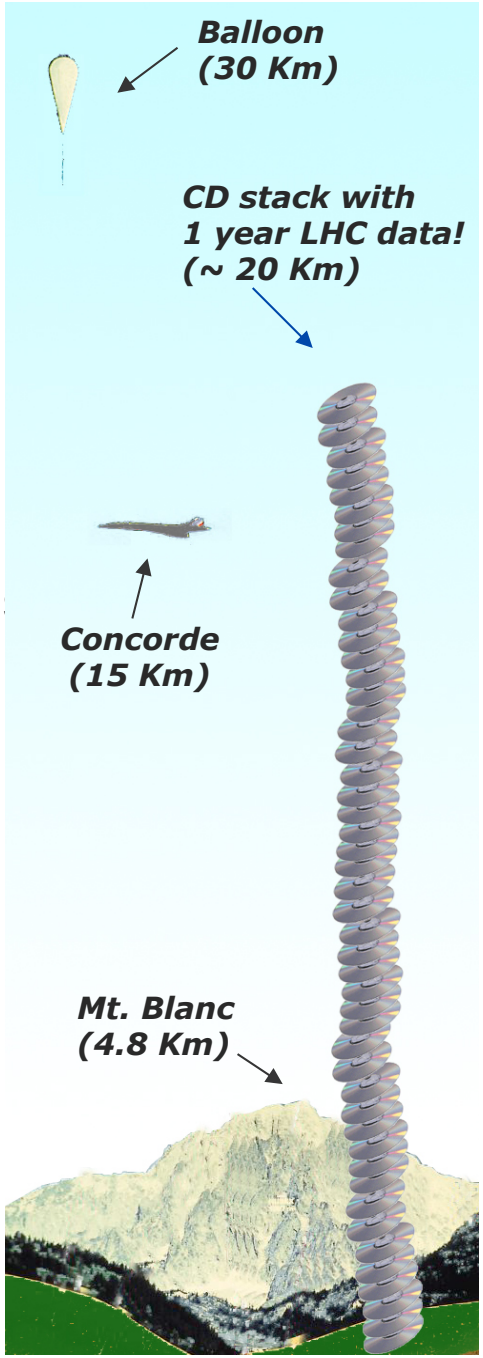
Antun Balaz
Scientific Computing Laboratory
Institute of Physics Belgrade, Serbia
http://www.scl.rs/

SEE-GRID-SCI
SEE-GRID eInfrastructure for regional eScience

30 Nov 2009 – 11 Dec 2009

Motivation

- **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 collaboration
  - e.g. Large Hadron Collider (LHC) at CERN
    - 40 million collisions per second
    - ~10 petabytes/year (~10 Million GBytes)

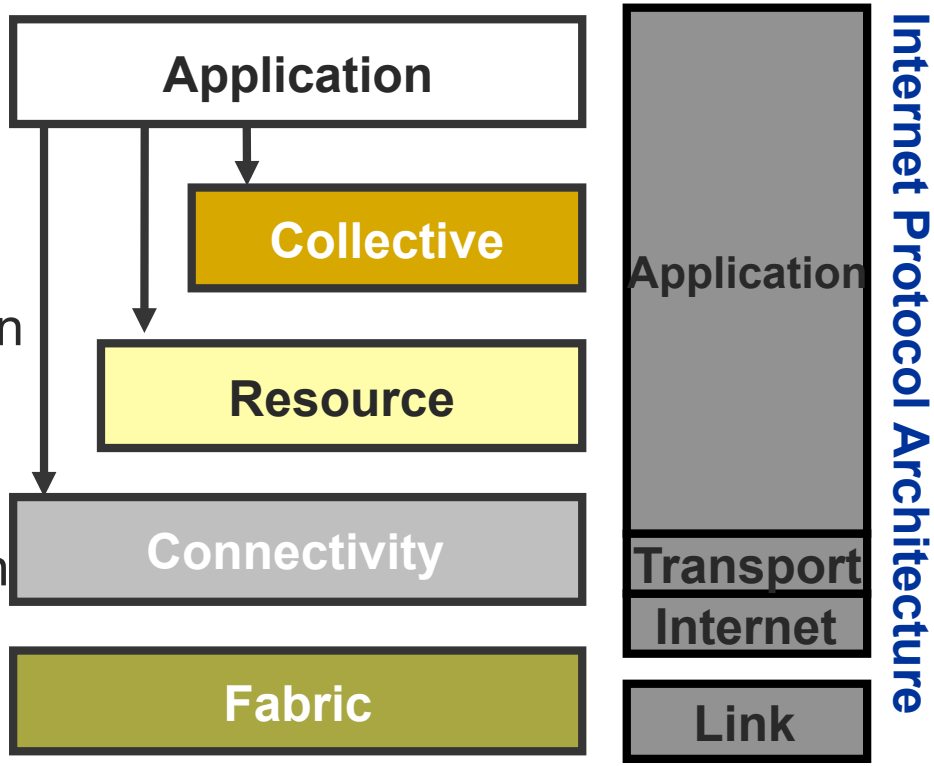


The solution: The Grid

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

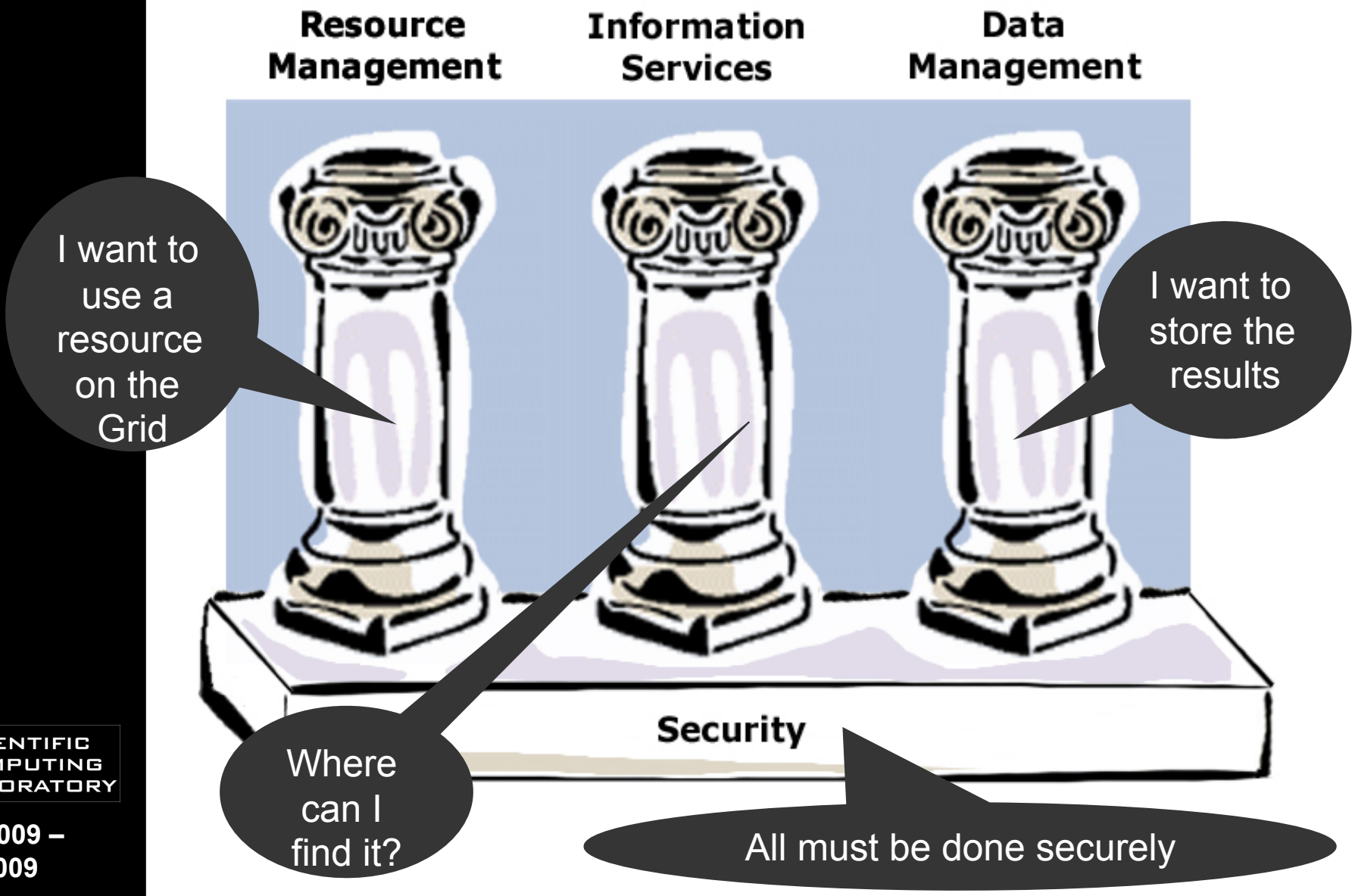
The Grid stack

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



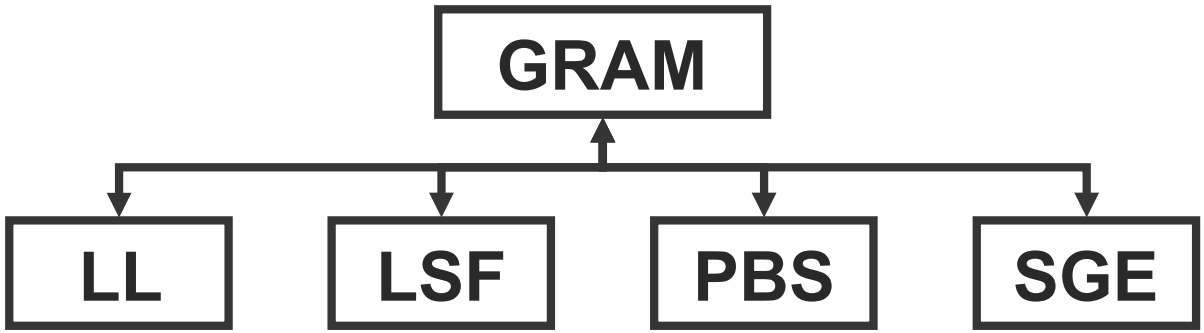
Grid foundations

- Defined by the Globus: http://globus.org



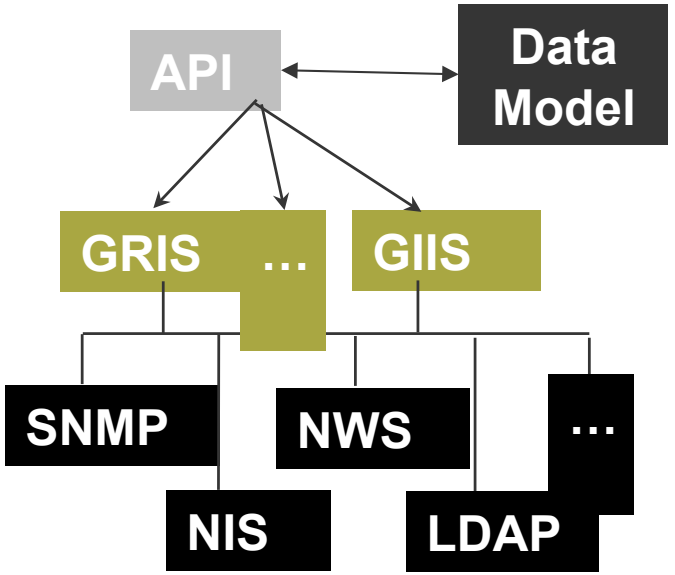
Resource Management

- 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



Information Services

- 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...”

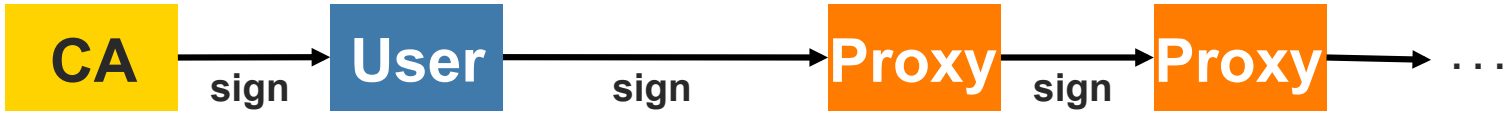


Data Management

- 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**

Security

- 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 PKI X.509 standard
    - CA signs certificates. Trust relationship
  - **Proxy certificates:** Temporary self-signed certs, allowing single sign-on: Proxy delegation



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 to the scientist

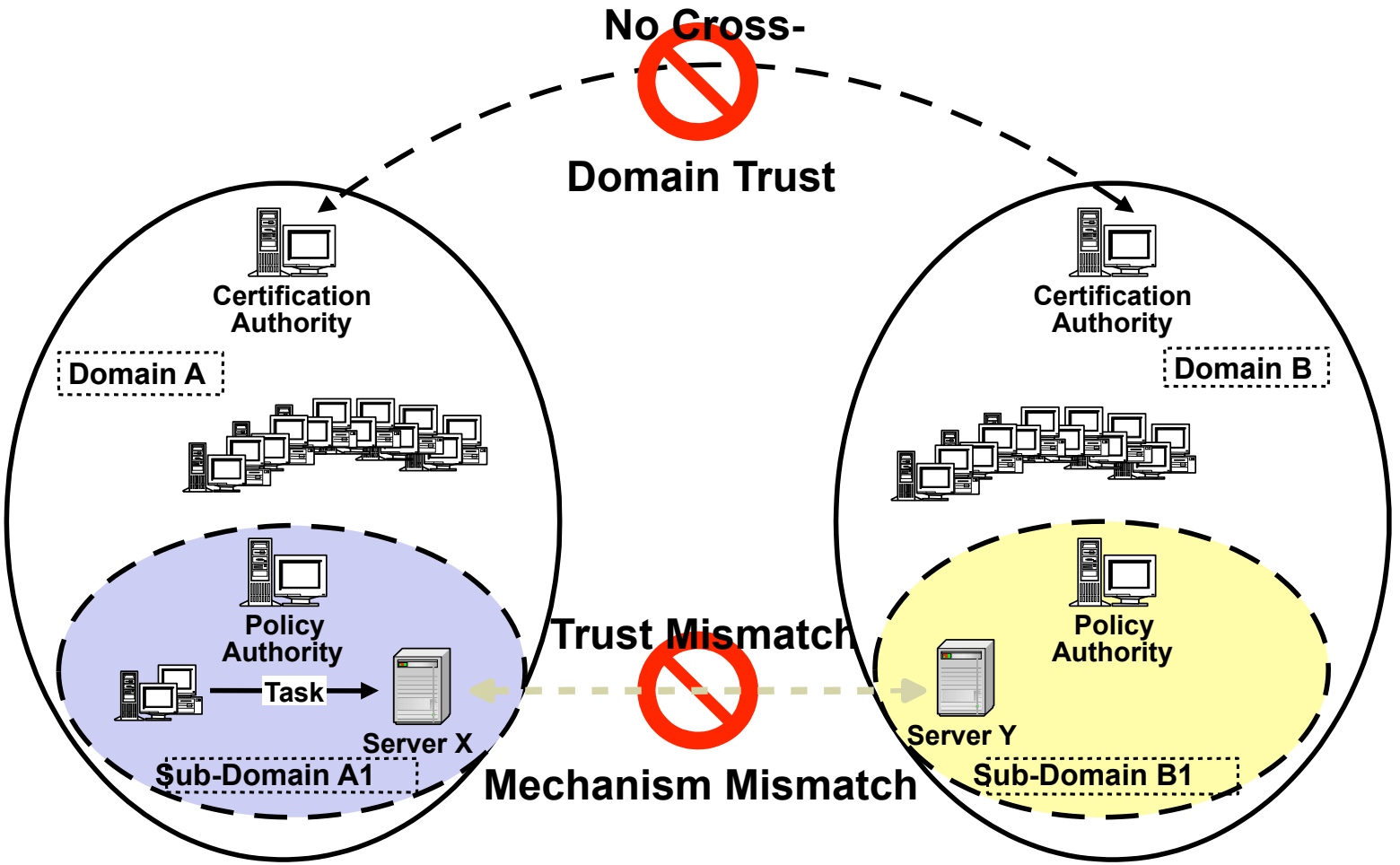
gLite – **Overview**

- First release 2005
- currently gLite 3.1-3.2
- 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

Set of basic Grid services

- Job submission/management
- File transfer (individual, queued database access)
- Data management (replication, metadata)
- Monitoring/Indexing system information

Multi-institution issues



SCIENTIFIC COMPUTING LABORATORY

30 Nov 2009 –
11 Dec 2009

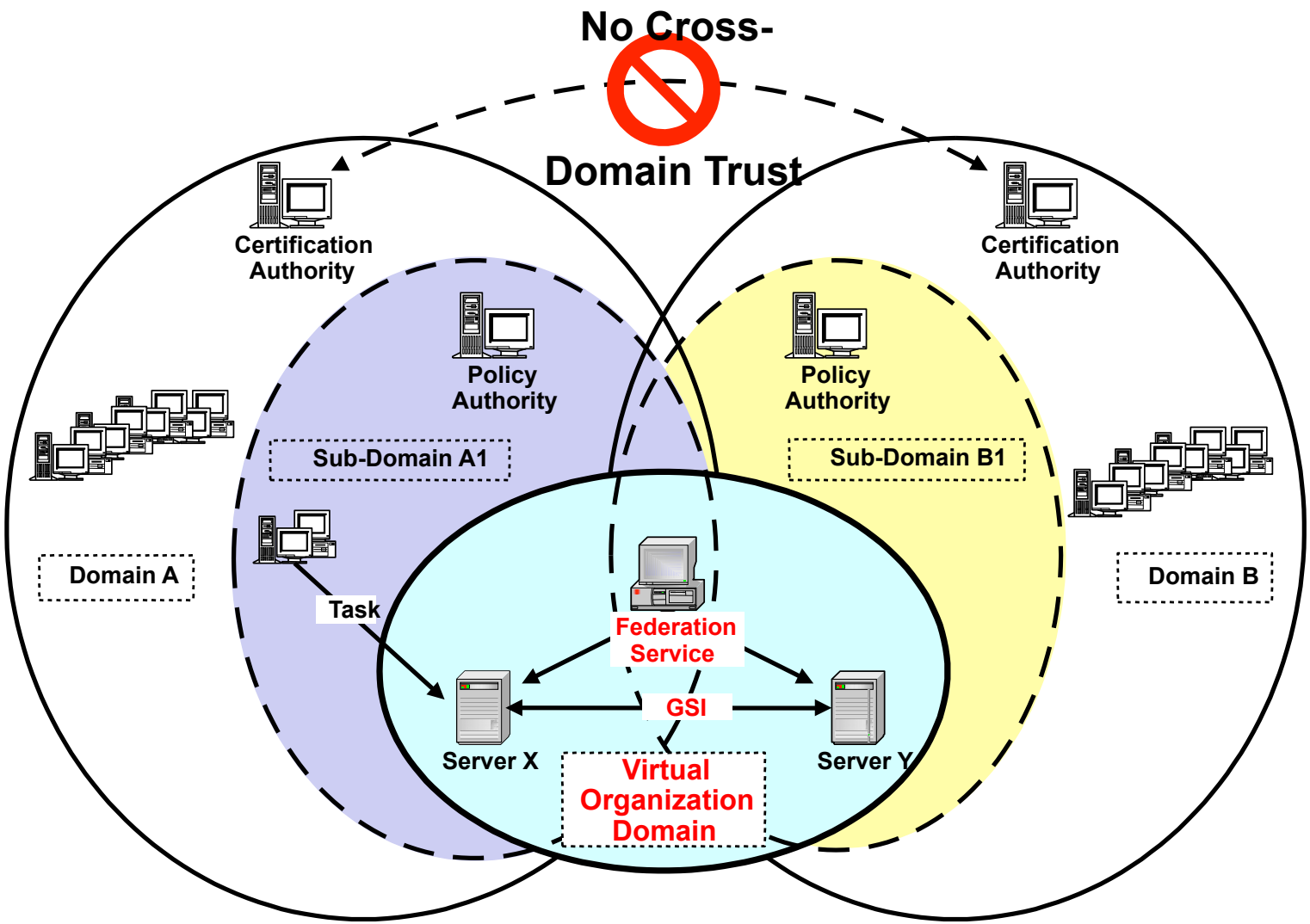
Why Grid security is hard (1)

- Resources being used may be valuable & the problems being solved sensitive
  - Both users and resources need to be careful
- Dynamic formation and management of user groups
  - Large, dynamic, unpredictable…
- Resources and users are often located in distinct administrative domains
  - Cannot assume cross-organizational trust agreements
  - Different mechanisms & credentials

Why Grid security is hard (2)

- Interactions are not just client/server, but service-to-service on behalf of user
  - Requires delegation of rights user ➔ service
  - Services may be dynamically instantiated
- Standardization of interfaces to allow for discovery, negotiation and use
- Implementation must be broadly available & applicable
  - Standard, well-tested, well-understood protocols; integrated with wide variety of tools
- Policy from sites, user communities and users need to be combined
  - Varying formats
- Want to hide as much as possible from applications!

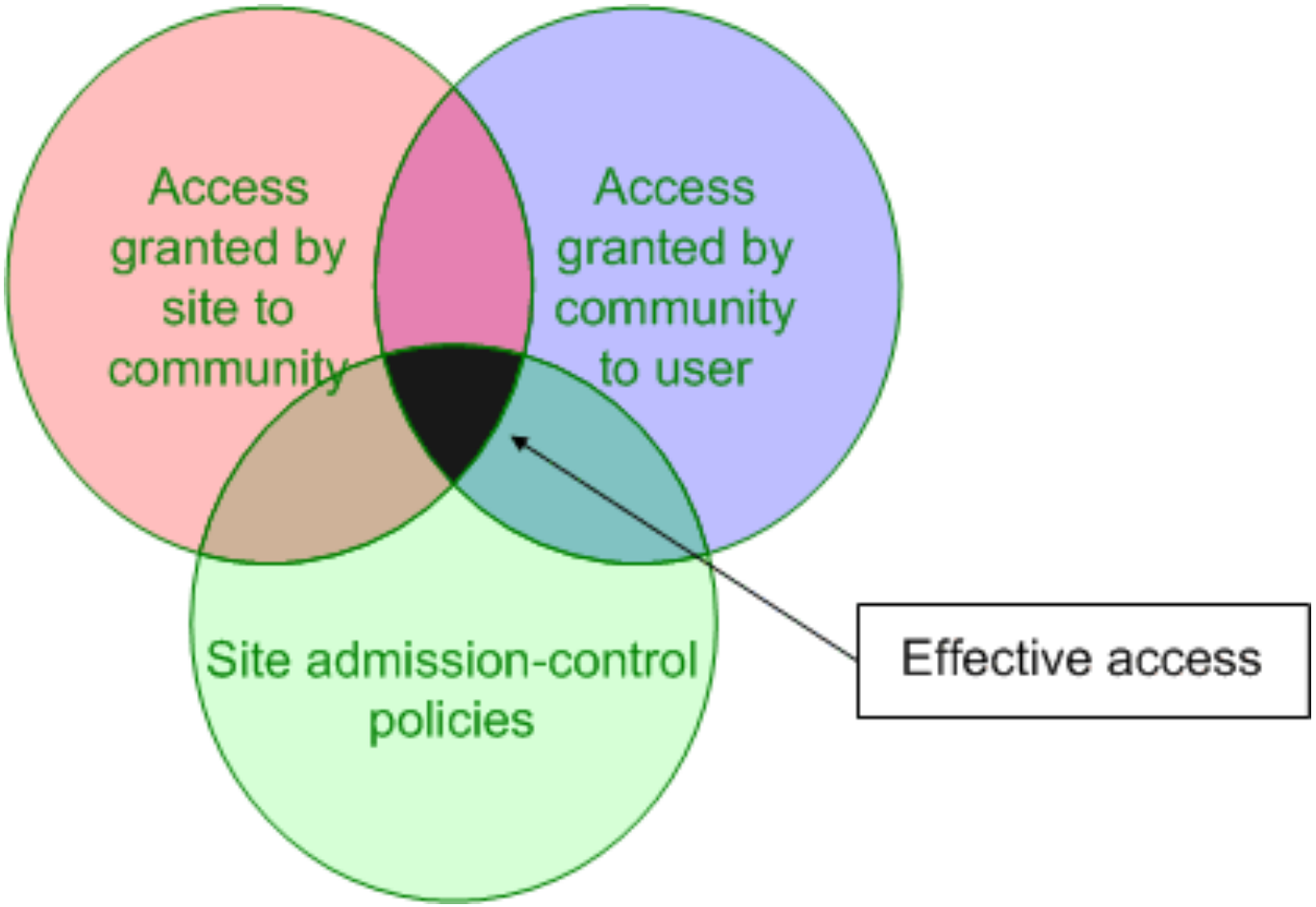
Grid solution: use of VOs



SCIENTIFIC COMPUTING LABORATORY

30 Nov 2009 – 11 Dec 2009

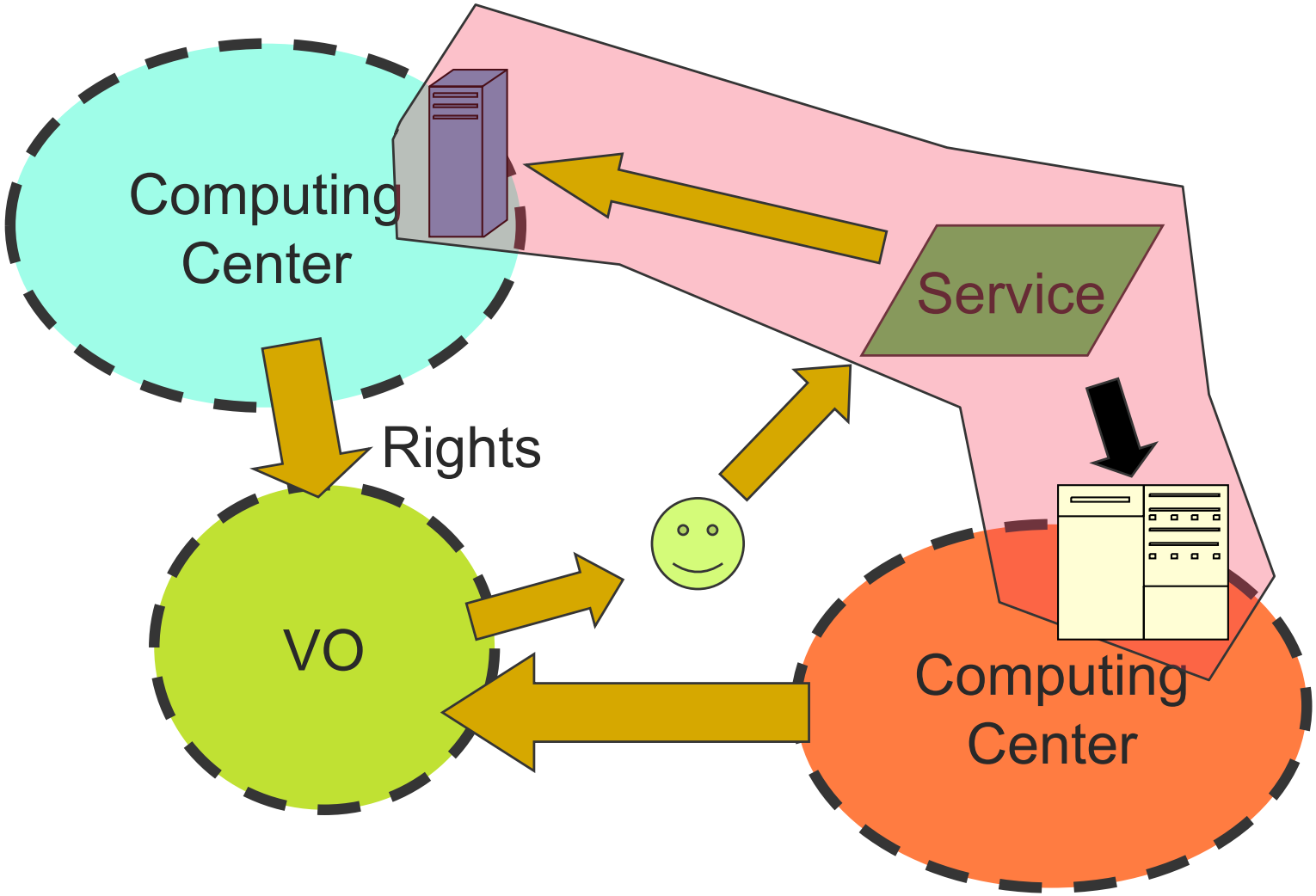
Effective policy governing access within a collaboration



Scientific Computing Laboratory

30 Nov 2009 – 11 Dec 2009

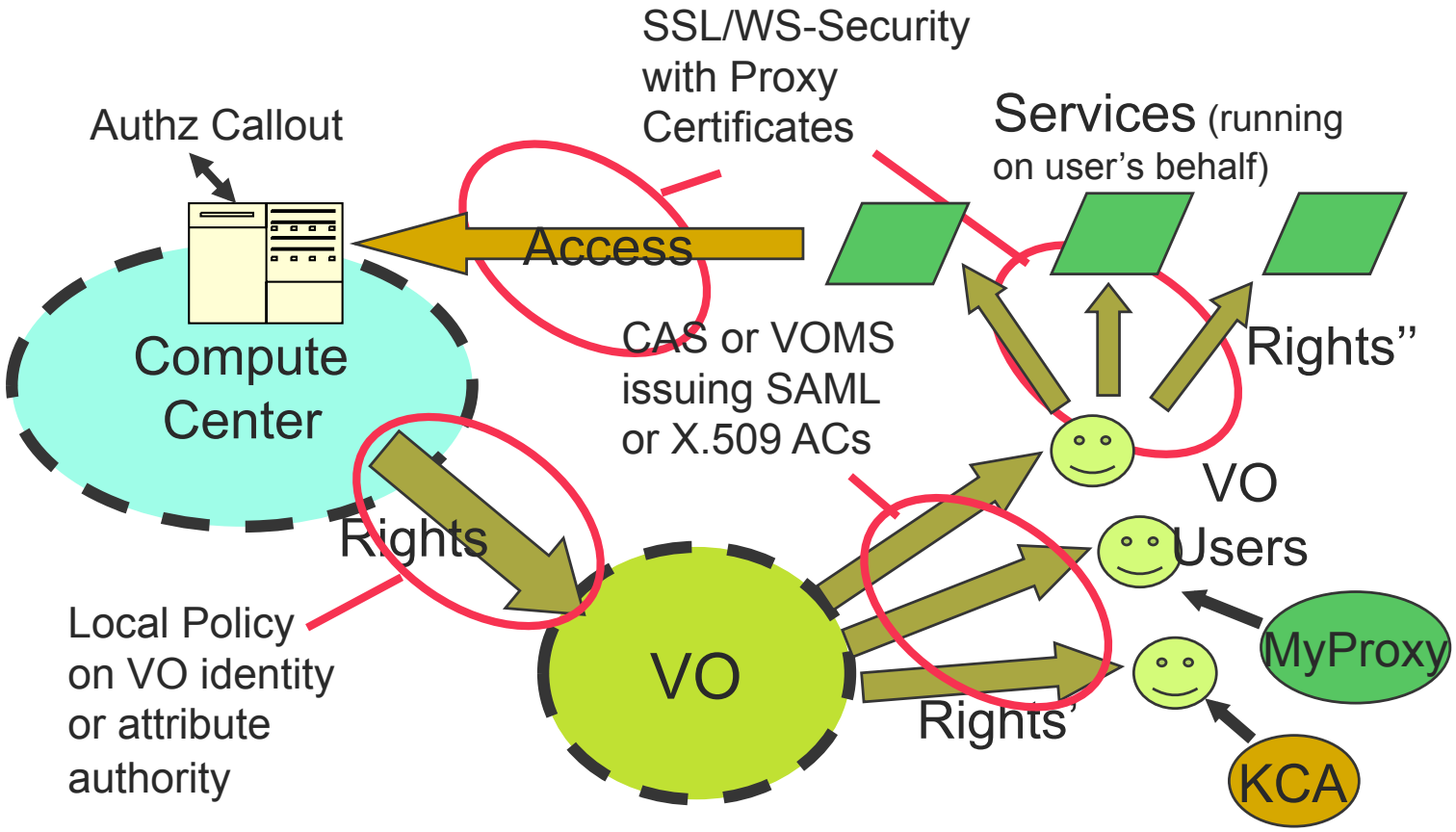
Use delegation to establish dynamic distributed system



SCIENTIFIC COMPUTING LABORATORY

30 Nov 2009 – 11 Dec 2009

GSI implementation



Grids and VOs (1)

- Virtual organizations (VOs) are groups of Grid users (authenticated through digital certificates)
- VO Management Service (VOMS) serves as a central repository for user authorization information, providing support for sorting users into a general group hierarchy, keeping track of their roles,etc.
- VO Manager, according to VO policies and rules, authorizes authenticated users to become VO members

Grids and VOs (2)

- Resource centers (RCs) may support one or more VOs, and this is how users are authorized to use computing, storage and other Grid resources
- VOMS allows flexible approach to A&A on the Grid

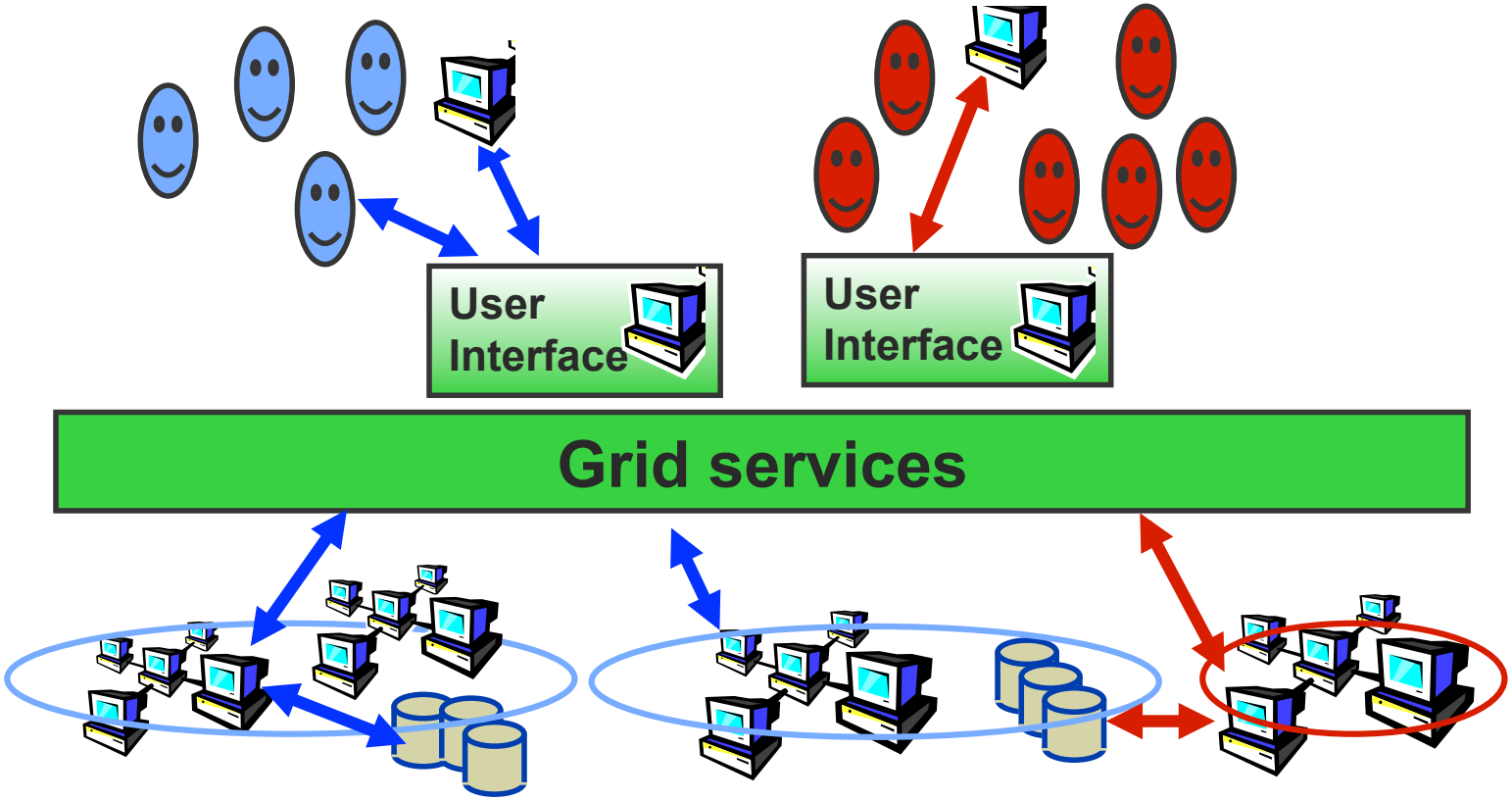
“Logging on” to the Grid

- To run programs, authenticate to Grid:

  voms-proxy-init –voms VONAME

  Enter PEM pass phrase: **********
- Creates a temporary, local, short-lived proxy credential for use by our computations
- Delegation = remote creation of a (second level) proxy credential, which allows remote process to authenticate on behalf of the user

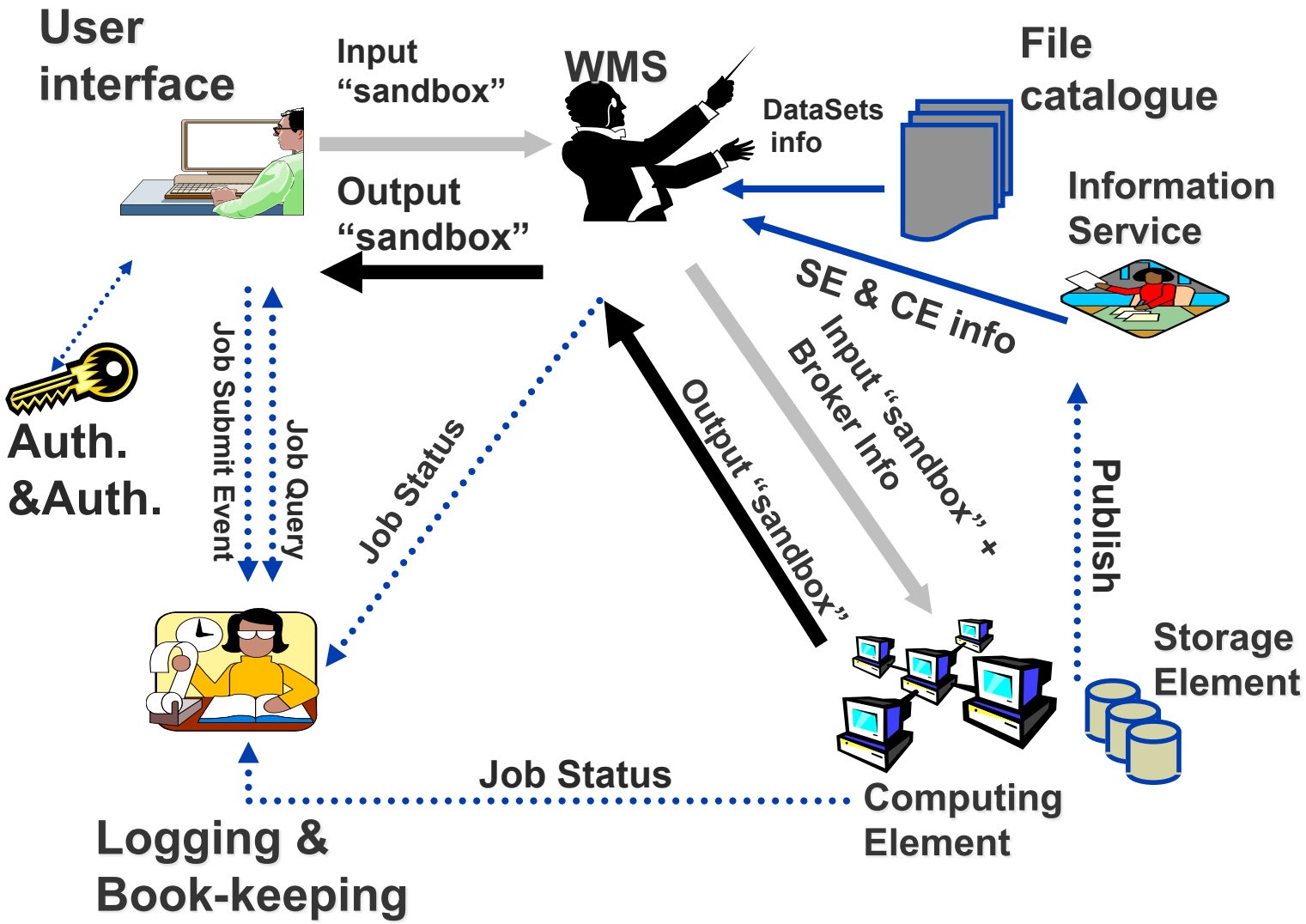
User view of the Grid



SCIENTIFIC COMPUTING LABORATORY

30 Nov 2009 – 11 Dec 2009

What really happens



SCIENTIFIC COMPUTING LABORATORY

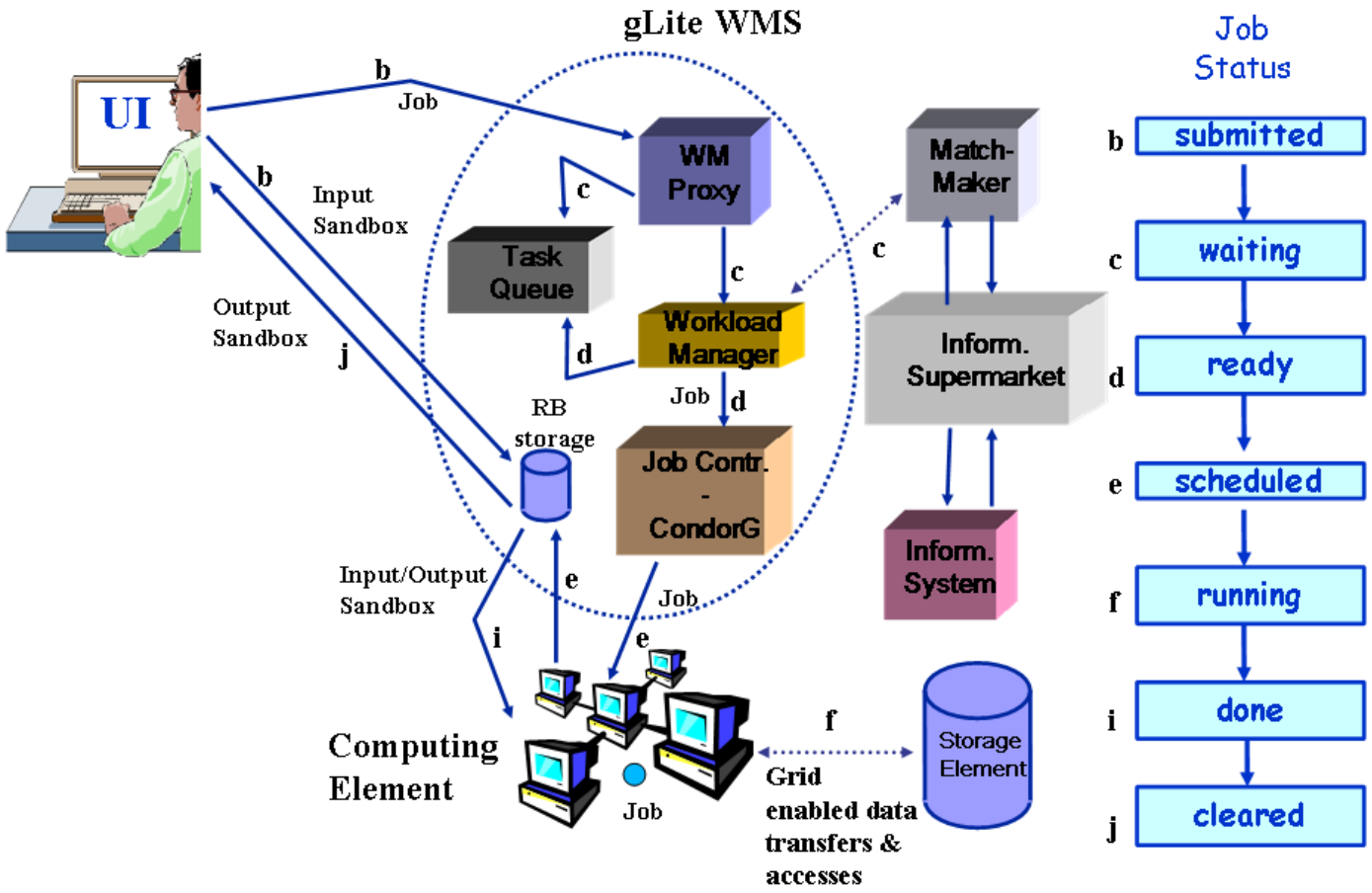
30 Nov 2009 – 11 Dec 2009

Workload Management System (WMS)

- Distributed scheduling
  - multiple UIs where you can submit your job
  - multiple WMSs from where the job can be sent to a CE
  - multiple CEs where the job can be put in a queuing system
- Distributed resource management
  - multiple information systems that monitor the state of the grid
  - Information from SE, CE, sites

SCIENTIFIC COMPUTING LABORATORY

WMS and job states



Authentication and Authorization

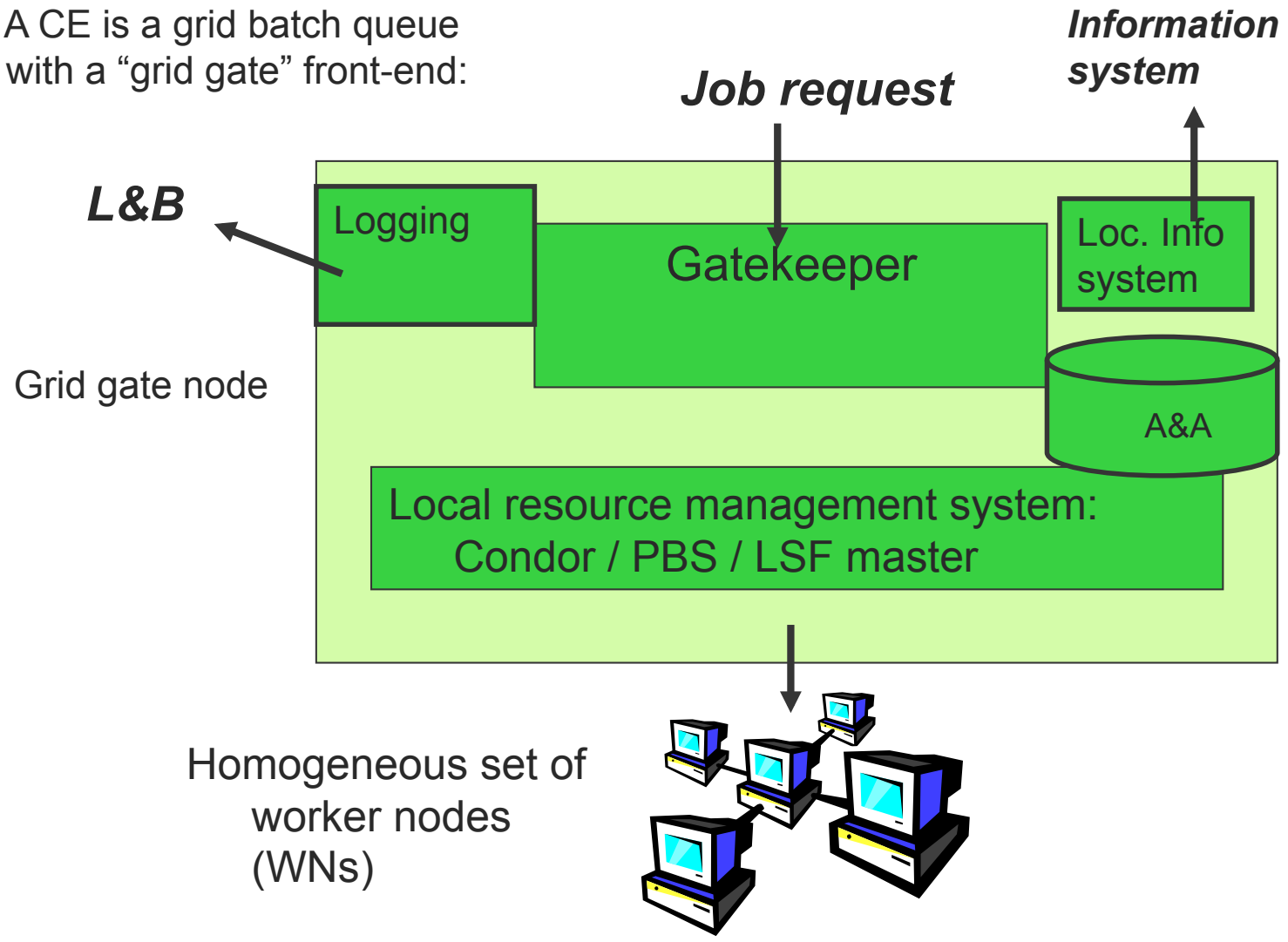
- Authentication
  - User obtains certificate from CA
  - Connects to UI by ssh
  - Downloads certificate
  - Invokes Proxy certificate
  - Single logon – to UI - then Secure Socket Layer with proxy identifies user to other nodes
- Authorization - currently
  - User joins Virtual Organisation
  - VO negotiates access to Grid nodes and resources (CE, SE)
  - Authorization tested by CE, SE: VOMS (or grid-mapfile) maps user to local accounts

User Interface (UI)

- UI is the user's interface to the Grid - Command-line interface to
  - Proxy certificate
  - Job operations
    - To submit a job
    - Monitor its status
    - Retrieve output
  - Data operations
    - Upload file to SE
    - Create replica
    - Discover replicas
  - Other grid services
- To run a job user creates a JDL (Job Description Language) file

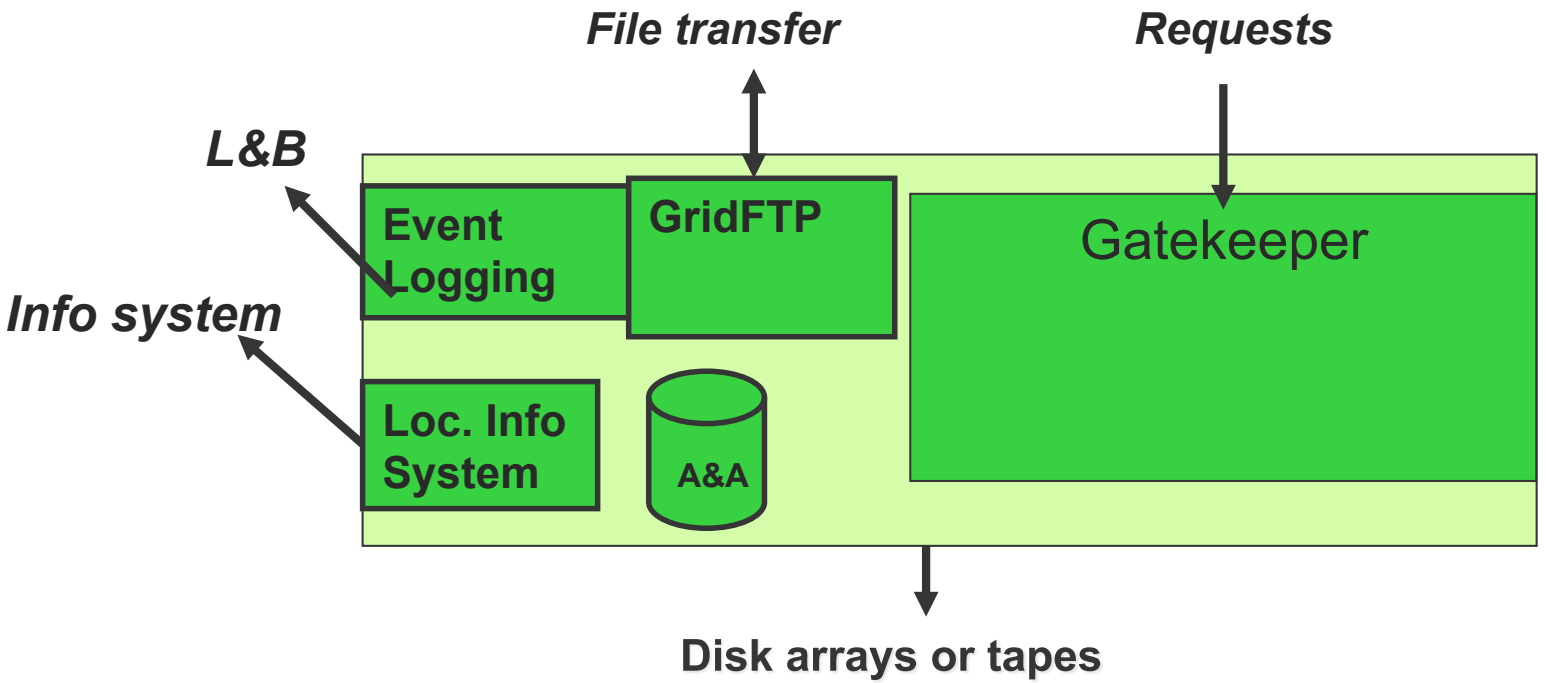
Computing Element (CE)

A CE is a grid batch queue with a “grid gate” front-end:



Storage Element (SE)

- Storage elements hold files: write once, read many
- Replica files can be held on different SE:
  - “close” to CE; share load on SE
- File Catalogue - what replicas exist for a file and where are they?



SCIENTIFIC COMPUTING LABORATORY

30 Nov 2009 – 11 Dec 2009

Logging and Bookkeeping

- Who did what and when?
- What is happening to my job?
- Usually runs on the WMS node

Information System

- Receives periodic (~5 min) updates from CE, SE, etc.
- Used by WMS (RB) node to determine resources to be used by a job
- Currently BDII is used

Typical Grid site

- CE + batch system + set of WNs
- SE + set of disk nodes
- MON: accounting and R-GMA
- BDII_site: collects information about all elements
- Additional services (WMS+LB, PX, VOMS, etc.)

Grid in a nutshell

- Grid structure is complicated but hidden from end-users, enabling all the comfort they need
- Users just need to join the VO and obtain certificates: we already have some VOs at hand for you!
- Use of Grid is then just as easy as the use of a typical Linux cluster