An overview of the EGEE infrastructure and middleware

Emmanuel Medernach

Based on a talk from Flavia Donno

EGEE is funded by the European Union under contract IST-2003-508833

Goals

- To introduce the major components of the EGEE grid
  - Architecture
  - Middleware
  - Organisation

Overview

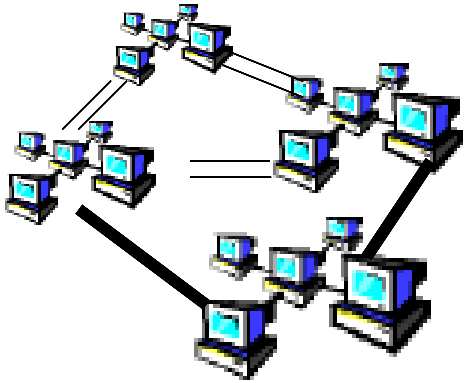
- Enabling Grid Computing:
  architecture + middleware + infrastructure
  - Authentication and Authorization
  - Information services
  - The major components of the infrastructure
  - The software stack
- EGEE grid organisation

What are the characteristics of a Grid system?

Numerous Resources

Ownership by Mutually Distrustful Organizations & Individuals

Connected by Heterogeneous, Multi-Level Networks



Different Security Requirements & Policies Required

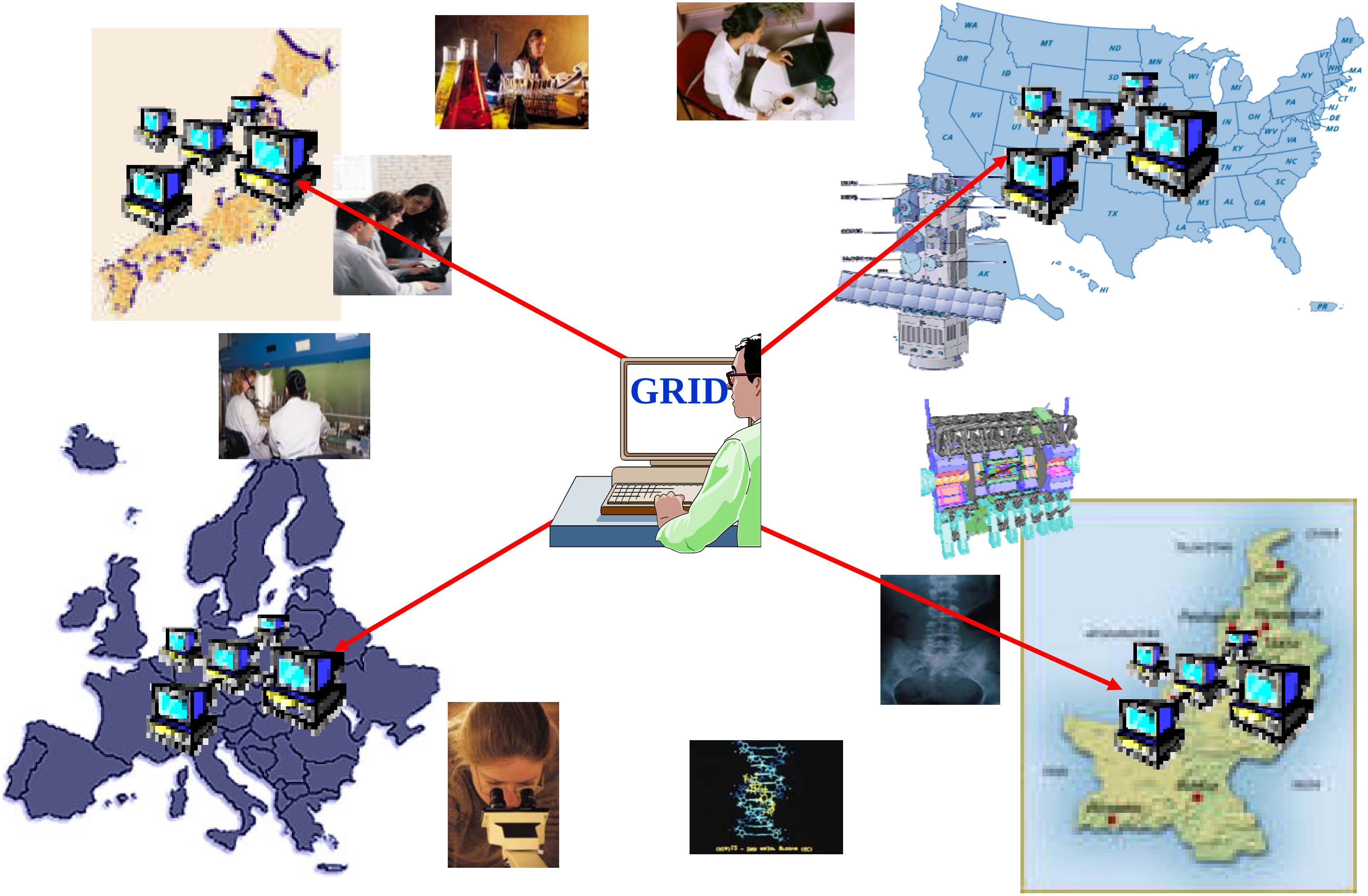
Different Resource Management Policies

Potentially Faulty Resources

Geographically Separated

Resources are Heterogeneous

What are the characteristics of a Grid system?



Virtual organisations

- An EGEE user must belong to a VO

- A VO
  - Controls access to specified resources
  - Usually comprises geographically distributed people
  - Requires the ability to know who has done what, and who will not be allowed to do it again…. Security.

- Current VO’s:
  - HEP communities, biology, astronomy,…

How do I login on the Grid ?

- Distribution of resources: secure access is a basic requirement
  - secure communication
  - security across organisational boundaries
  - single “sign-on" for users of the Grid
- Two basic concepts:
  - Authentication: *Who am I?*
    - “Equivalent” to a pass port, ID card etc.

  - Authorisation: *What can I do?*
    - Certain permissions, duties etc.

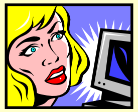
Security in the Grid

- In industry, several security standards exist:
  - Public Key Infrastructure (PKI)
    - PKI keys
    - SPKI keys (focus on authorisation rather than certificates)
    - RSA
  - Secure Socket Layer (SSL)
    - SSH keys
  - Kerberos
- Need for a common security standard for Grid services
  - Above standards do not meet all Grid requirements (e.g. delegation, single sign-on etc.)
- Grid community mainly uses X.509 PKI for the Internet
  - Well established and widely used (also for www, e-mail, etc.)

PKI – Basic overview

- Public Key Infrastructure (also called asymmetric cryptography)
- One primary advantage: it is generally easier than distributing secret keys securely, as required in symmetric keys

Entity A (Alice)



public key e
private key d

applies the decryption transformation

$$m = D_d(c).$$

Entity B (Bob)

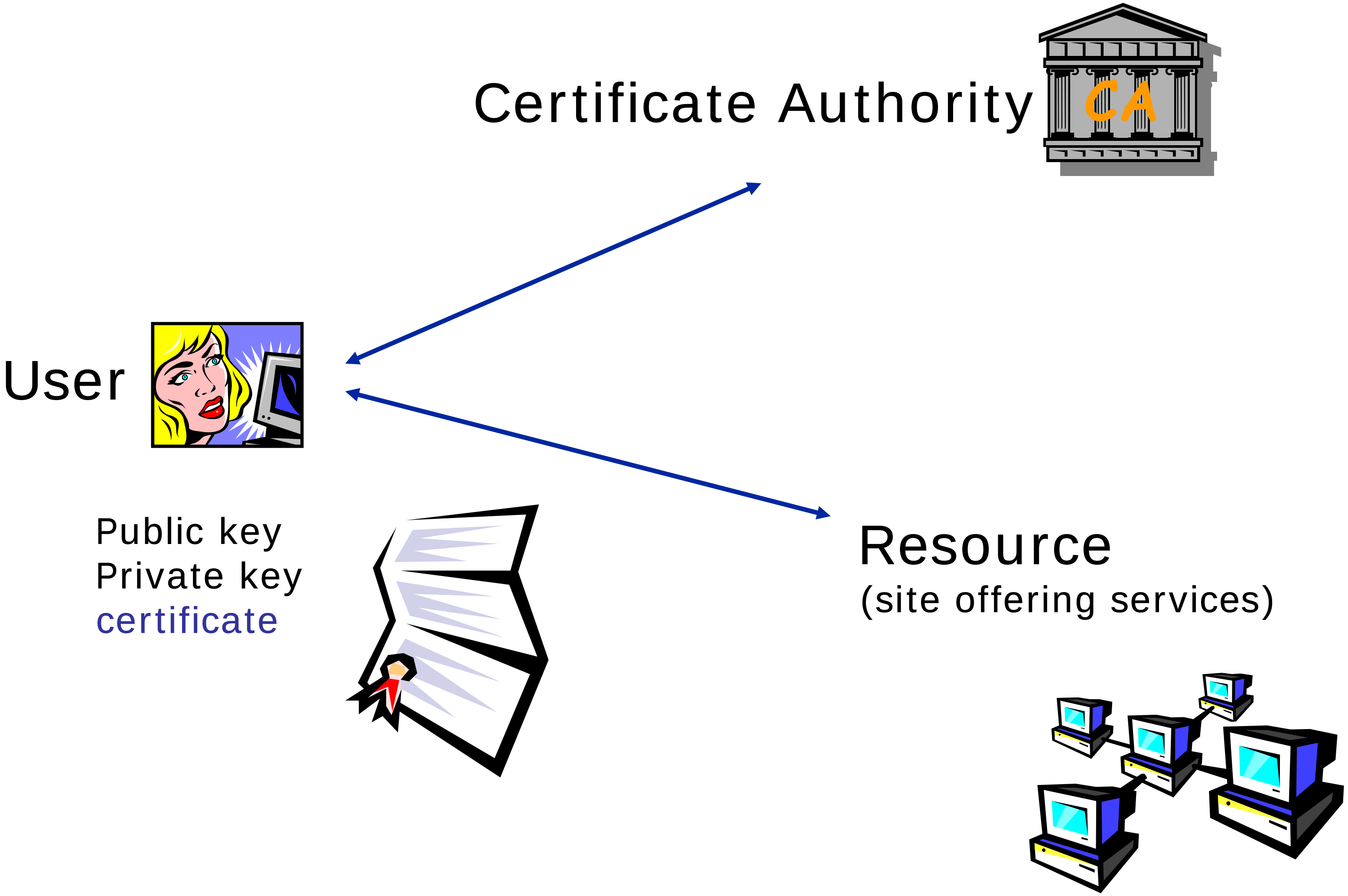
public key
private key

wishing to send a **message m** to A:

ciphertext $c = E_e(m)$

encryption transformation E_e
decryption transformation D_d

Involved entities



Certificate Authority

User

Public key
Private key
certificate

Resource
(site offering services)

X.509 alias ISO/IEC/ITU 9594-9

- X.509 certificate includes:
  - User identification (someone’s subject name)
  - Public key
  - A “signature” from a Certificate Authority (CA) that:
    - Proves that the certificate came from the CA.
    - Warranty for the subject name
    - Warranty for the binding of the public key to the subject

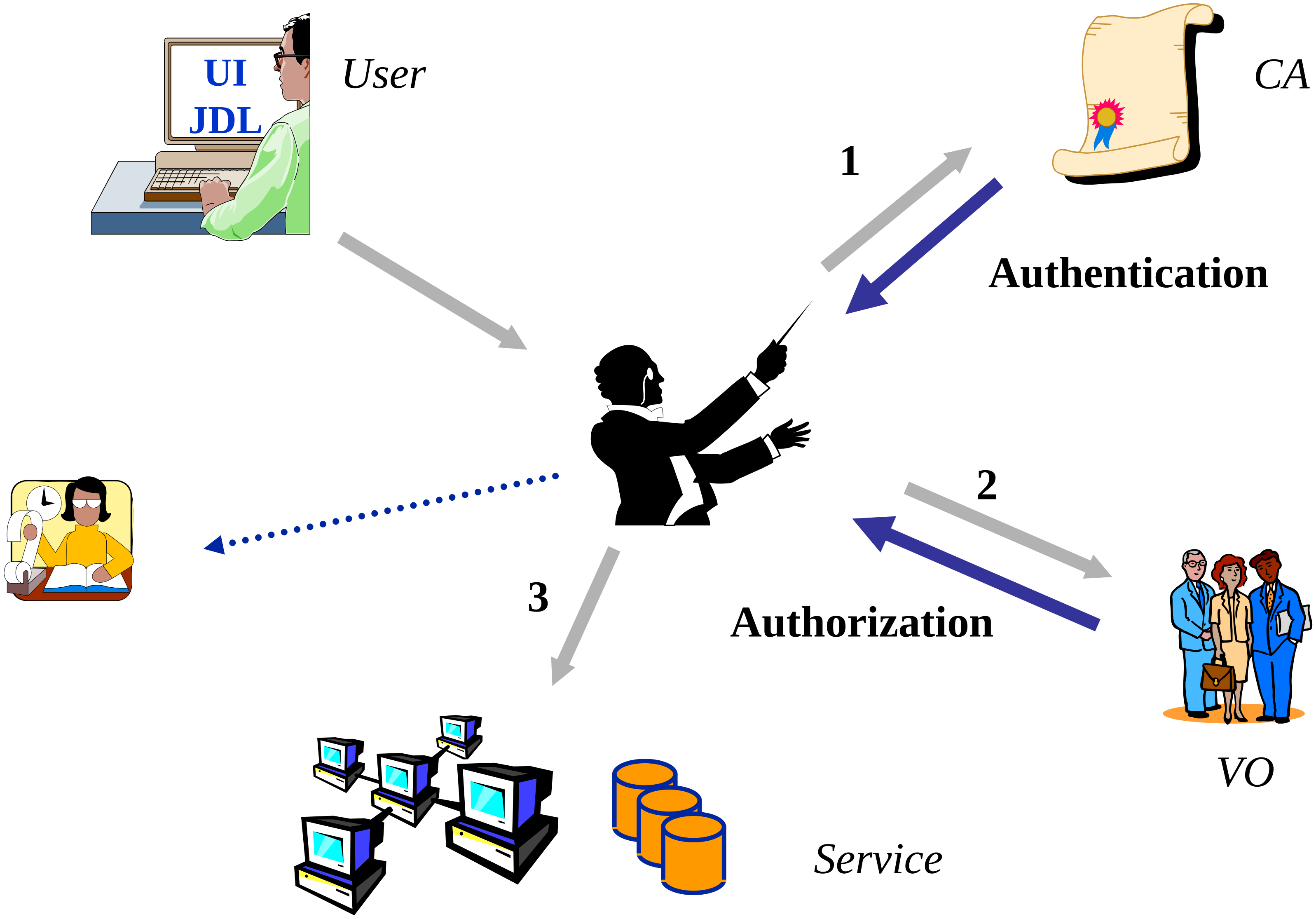
Grid Security Infrastructure (GSI)

- Globus Toolkit^{TM} proposed and implements the Grid Security Infrastructure (GSI)
  - Protocols and APIs to address Grid security needs
- GSI protocols extend standard public key protocols
  - Standards: X.509 & SSL/TLS
  - Extensions: X.509 Proxy Certificates (single sign-on) & Delegation
- GSI extends standard GSS-API (Generic Security Service)
  - The GSS-API is the IETF standard for adding authentication, delegation, message integrity, and message confidentiality to applications.
- Proxy Certificate:
  - Short term, restricted certificate that is derived form a long-term X.509 certificate
  - Signed by the normal end entity cert, or by another proxy
  - Allows a process to act on behalf of a user
  - Not encrypted and thus needs to be securely managed by file system

Authorisation Requirements

- Detailed user rights need to be centrally managed and assigned
  - User can have certain group membership and roles
- Involved parties:
  - Resource providers (RP, provides access to the resource)
    - keep full control on access rights
    - traceability user level (not VO level)
  - Virtual Organisation (VO) of the user (member of a certain group should have same access rights independent of resource)
- Agreement required between resource providers and VO
  - RPs evaluate authorisation granted by VO to a user and map into local credentials to access resources
- Need tool to manage membership for large VOs (10,000 users)

Authentication and Authorization



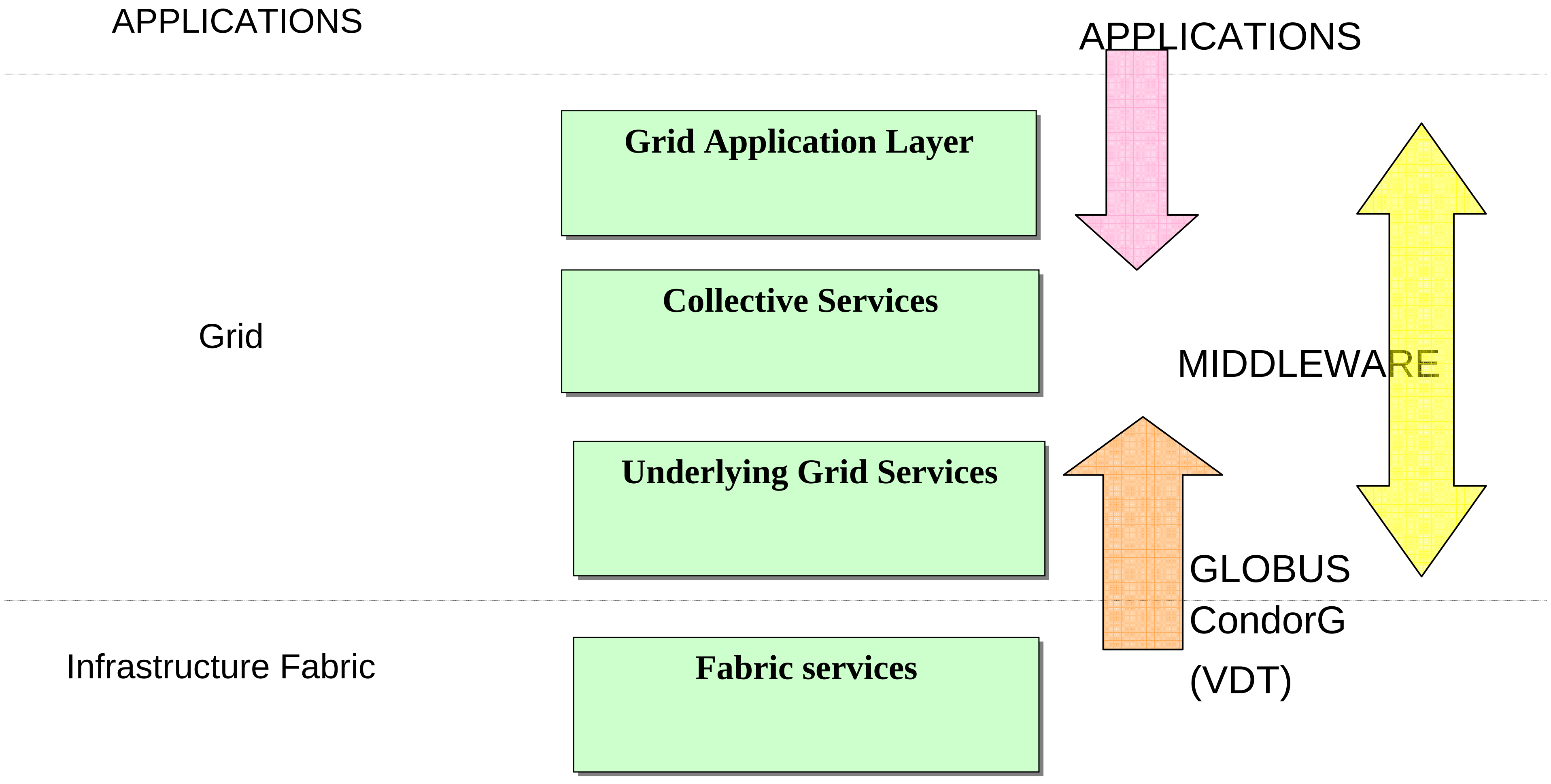
Security Summary

- Security is important for Grid middleware:
  - In particular in commercial use
- Security solutions need to be integrated from the very beginning
- Grid security relies on PKI
  - Requires: authentication & authorization
- Basic entities:
  - Users – CA (Certificate Authorities) – Resource Providers

The middleware

- EGEE middleware built upon toolkits provides generic Grid services:
  - Information
  - Job submission
  - Data management
  - Security
  - Logging
  - Monitoring
- EGEE supports computation and data storage by multiple virtual organisations

The grid software stack



APPLICATIONS

Grid

Infrastructure Fabric

Grid Application Layer

Collective Services

Underlying Grid Services

Fabric services

APPLICATIONS

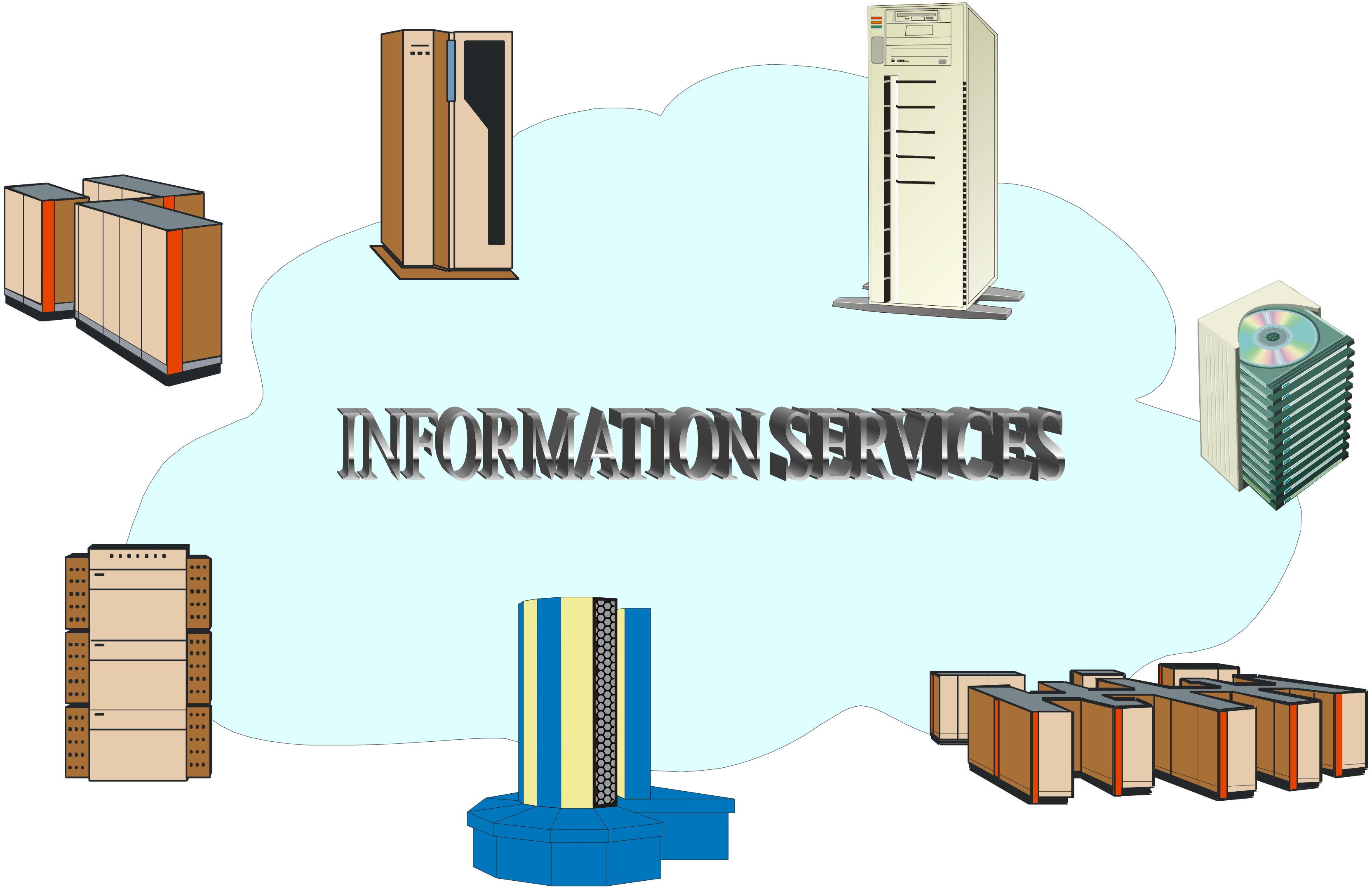
MIDDLEWARE

GLOBUS

CondorG

(VDT)

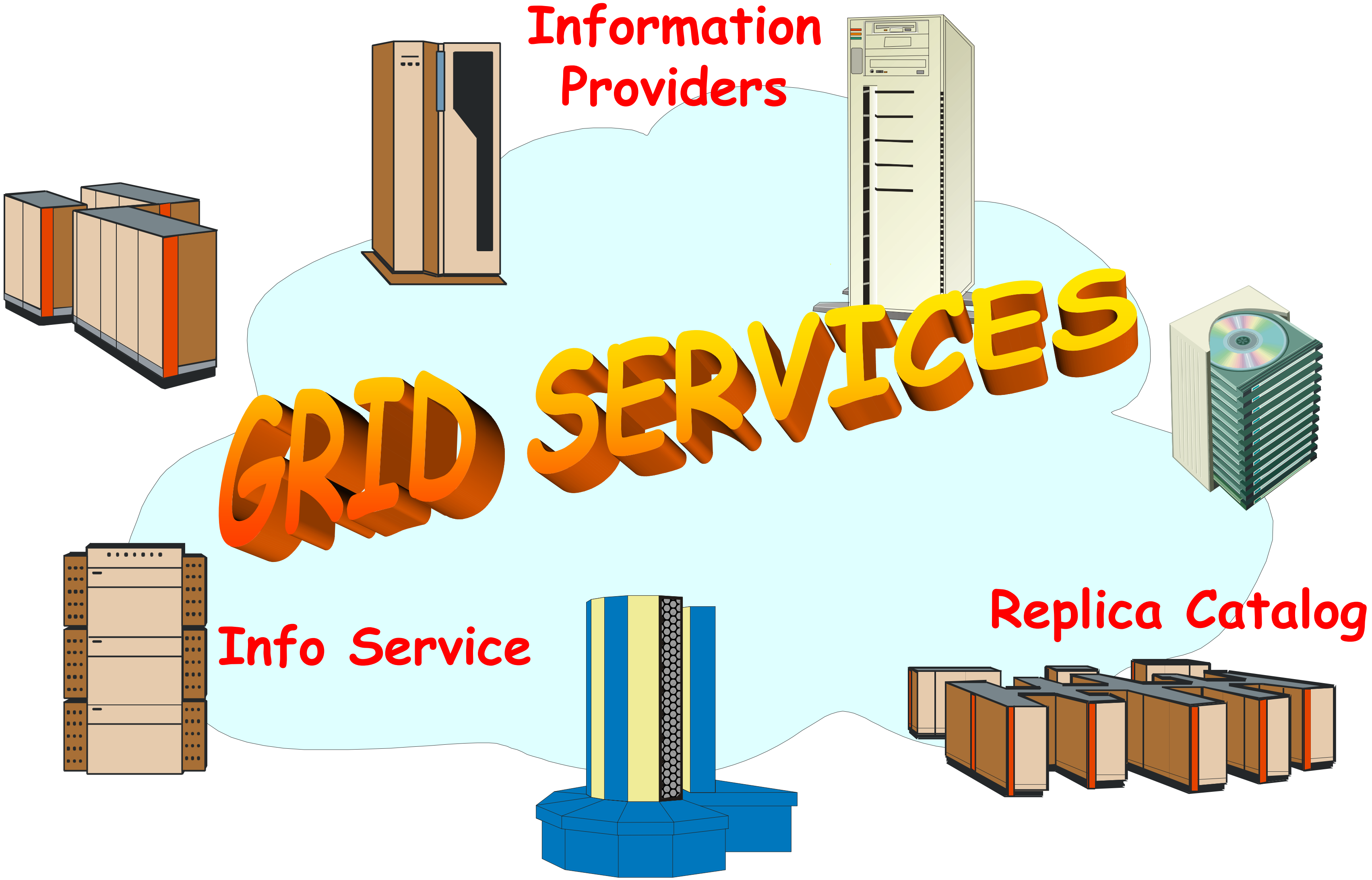
Situation on a Grid



Features of a grid information system

- Provides information on both:
  - The Grid itself
    - Mainly for the middleware packages
    - The user may query it to understand the status of the Grid
  - Grid applications
    - For users
- Flexible architecture
  - Able to cope with nodes in a distributed environment with an unreliable network
  - Dynamic addition and deletion of information producers
  - Security system able to address the access to information at a fine level of granularity
  - Allow new data types to be defined
  - Scaleable
  - Good performance
  - Standards based

Situation on a Grid



Information Services

- Hardware:
  - EDG Information Service
  - Information Providers
- Data:
  - Replica Catalog
    - LDAP (release 1.4)
    - RLS (release 2.0)
- Software & Services:
  - EDG Grid Services:
    - Information Service
      - MDS
      - R-GMA
  - Application Services:
    - Currently only EDG applications directly supported

Machine Types:

- Information Service (IS)
  - Top level MDS
  - R-GMA registry
- Replica Catalog (RC, RLS)

EDG Information Providers

EDG information providers

- Software that provides information about resources and infrastructure
- Provided by the developer of a service or the responsible for the resource

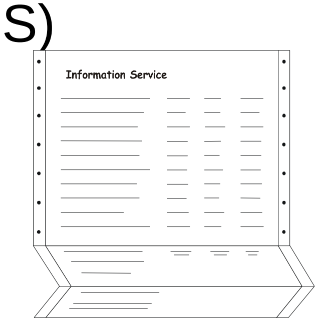
Main EDG Grid Services

- Authentication & Authorization
- Job submission service
  - Resource Broker
- Replica Management
  - EDG-Replica-Manager
  - Mass storage system support
- Monitoring

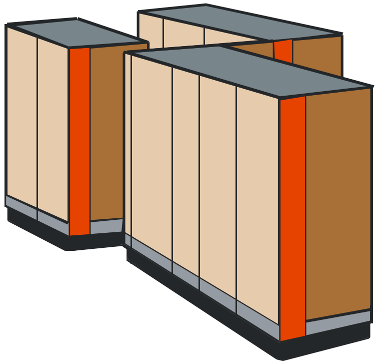
Main Logical Machine Types

- User Interface (UI)

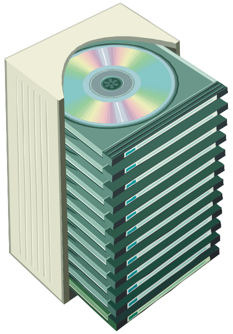
- Information Service (IS)



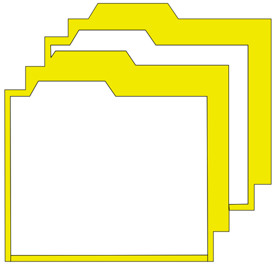
- Computing Element (CE)
  - Frontend Node
  - Worker Nodes (WN)



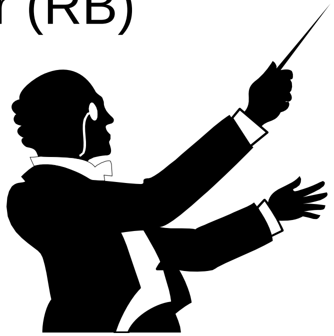
- Storage Element (SE)



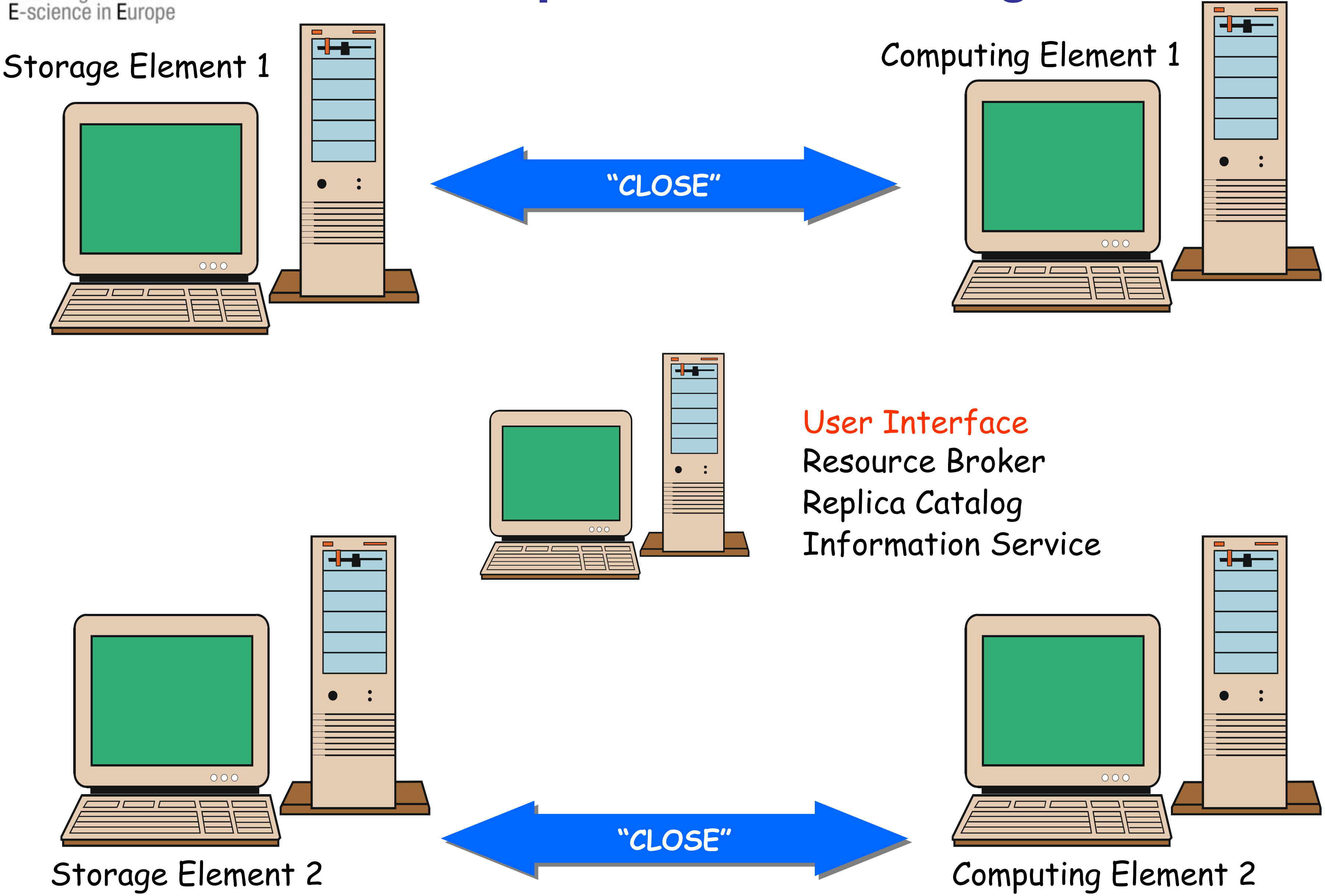
- Replica Catalog (RC,RLS)



- Resource Broker (RB)



A Simple Testbed Configuration



Storage Element 1

Computing Element 1

"CLOSE"

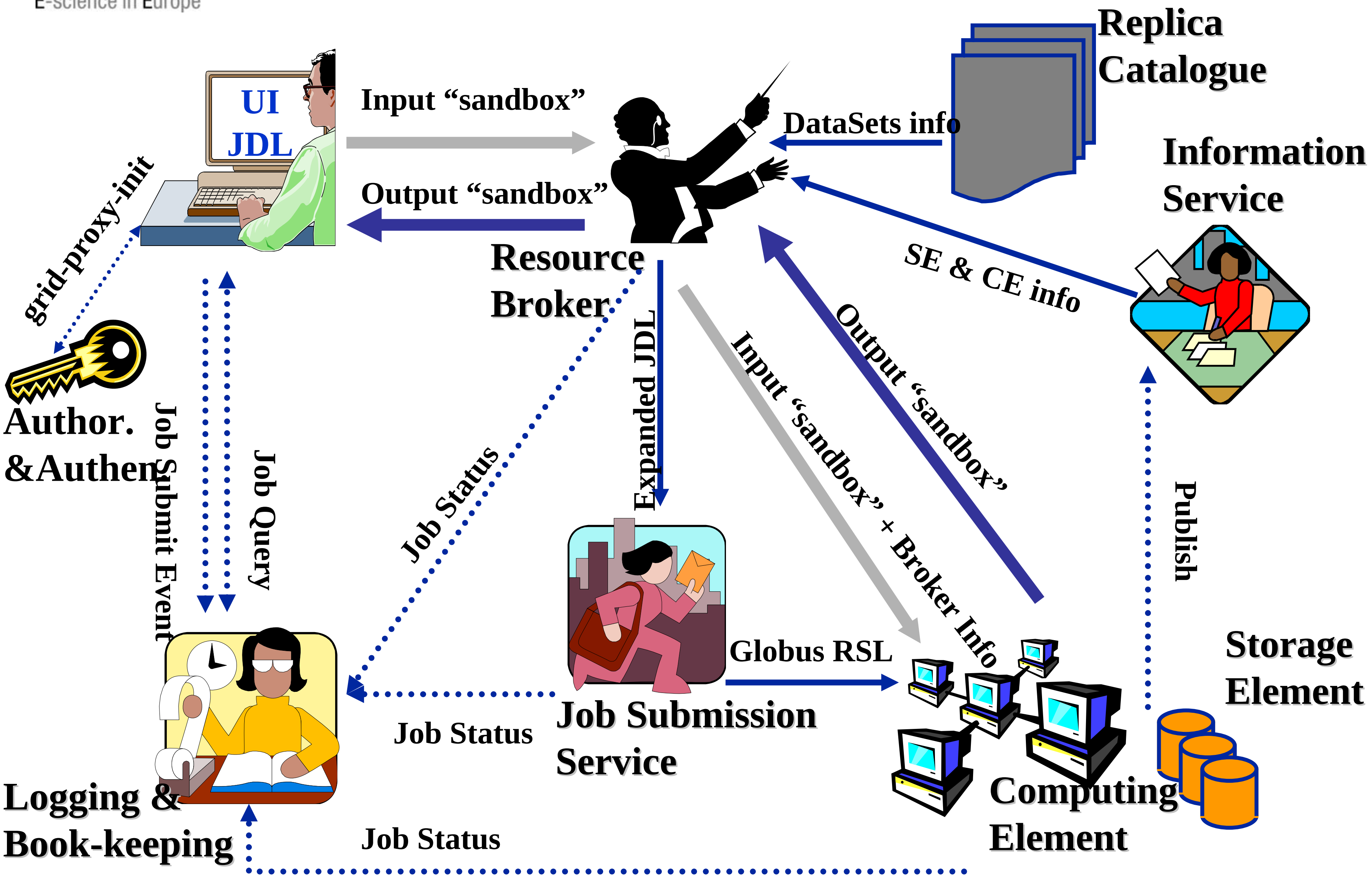
User Interface
Resource Broker
Replica Catalog
Information Service

Storage Element 2

"CLOSE"

Computing Element 2

The lifecycle of an EGEE job



Main Services per Machine Type

Daemon	UI	IS	CE (frontend)	WN	SE	RLS	RB
Globus Gatekeeper	-	-	✓	-	-	-	-
RLS-LRC	-	-	-	-	-	✓	-
RLS-RMC	-	-	-	-	-	✓	-
GridFTP	-	-	✓	-	✓	-	✓
R-GMA	-	✓	-	-	-	-	-
R-GMA GOUT	-	-	-	-	-	-	✓
R-GMA GIN	-	-	✓	-	✓	-	-
Broker (Network server, job control)	-	-	-	-	-	-	✓
CondorG Job submission	-	-	-	-	-	-	✓
Logging & Bookkeeping	-	-	-	-	-	-	✓
Local Logger	-	-	✓	-	-	-	✓
CRL Update	-	-	✓	-	✓	-	✓
Grid mapfile Update	-	-	✓	-	✓	-	✓
RFIO	-	-	-	-	✓	-	-
EDG-SE	-	-	-	-	✓	-	-

EGEE

- EGEE is distinctive because of the emphasis on :
  - Production quality of service
  - Multiple virtual organisations

Training/demo service

- Permanent need for tutorials, demonstrations etc.
- Cannot disturb production system, or guarantee pre-production
- Ideally need dedicated (small) service
  - Kept in an operational state
  - Need sufficient resources to be available (another testbed!)
- Currently fulfilled by GILDA service

Conclusions

- The EGEE Grid requires resources, an infrastructure and middleware that allows for:
  - Authentication and Authorization
  - Information services
  - Job and Data Management
  - Monitoring and fault recovery
- We have seen the main components of the EGEE Grid Service and Organization
  - EGEE is VO based
  - The Grid Operations Management Structure monitors and controls the overall functionality
- The EGEE tutorials ensure training at all levels with hands-on on the GILDA dedicated testbed