Introduction to EGEE and the GILDA t-infrastructure

Robin McConnell rmcconne@nesc.ac.uk
Training & Induction

Gergely Sipos sipos@sztaki.hu
Training and induction
Application Porting Support

www.eu-egee.org

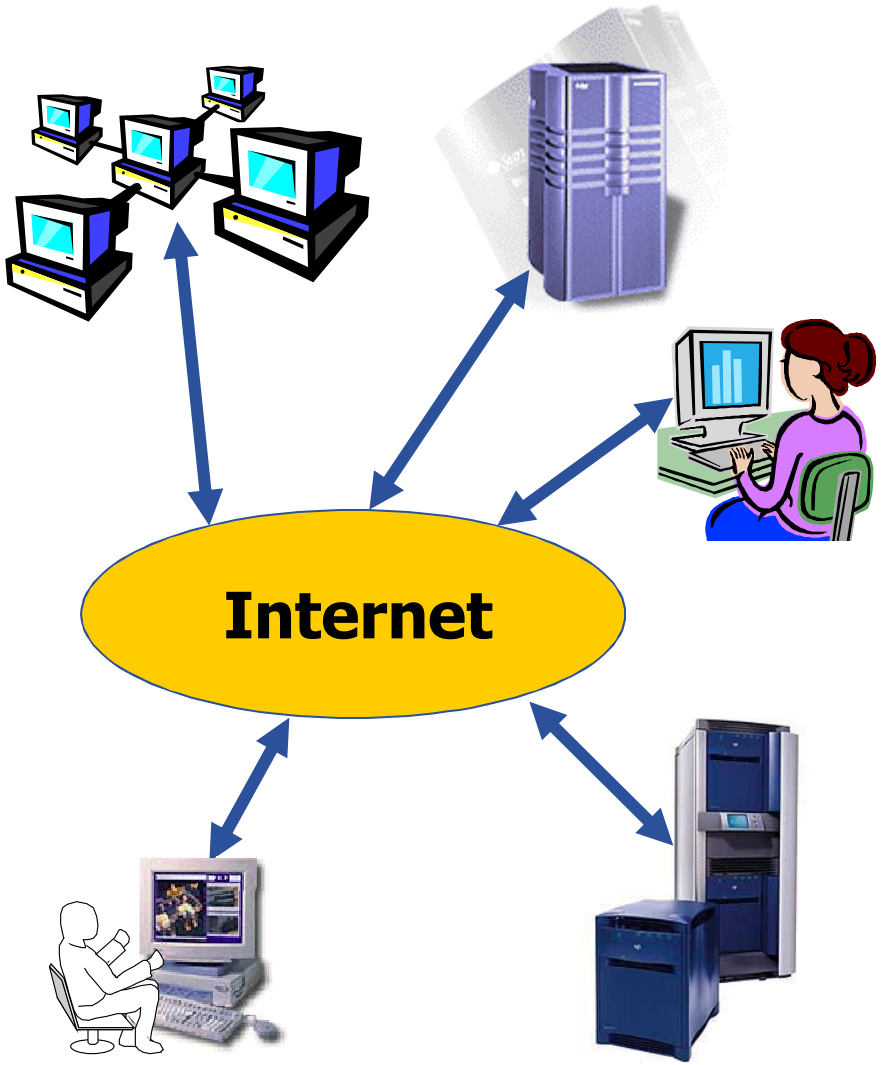
e-infrastructure

Outline

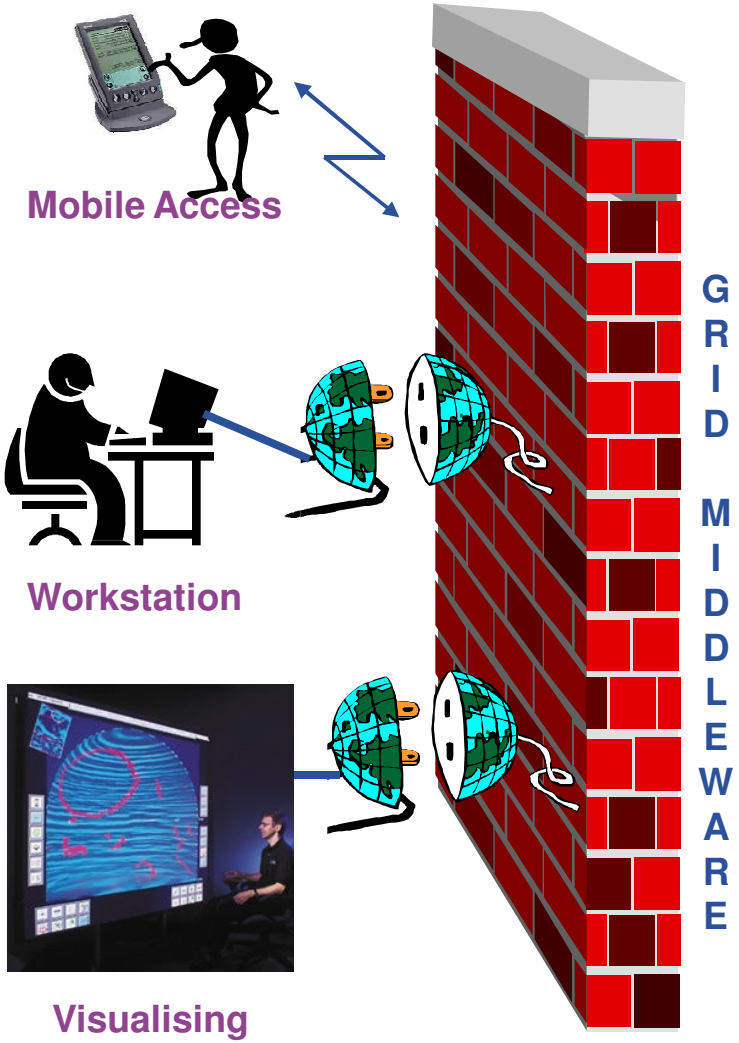
- **What is the grid?**
- **What is EGEE, how to get access**
  - To a general infrastructure
  - To GILDA training infrastructure
- **Communities, applications, examples**
- **User support services**

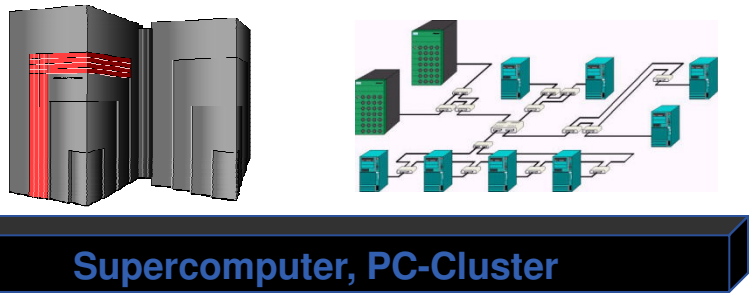
What is the Grid?

- A Grid is a collection of computers, storages, special devices, services that can **dynamically join and leave** the Grid
- They are **heterogeneous** in every aspect
- They are geographically **distributed** and connected by a **wide-area network**
- They can be accessed **on-demand** by a set of users in a controlled way

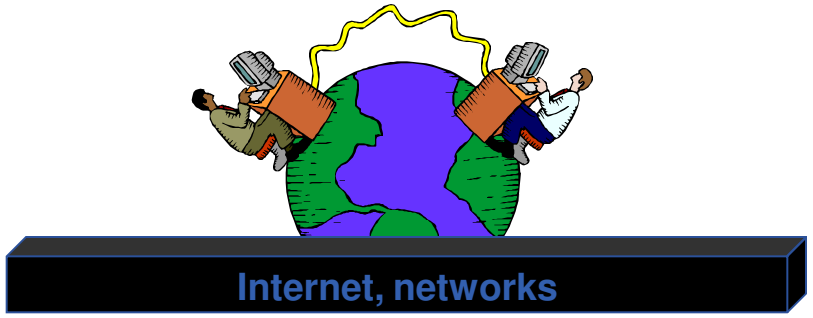


Grid vision


Mobile Access
Workstation
Visualising
GRID MIDDLEWARE


Supercomputer, PC-Cluster

Data-storage, Sensors, Experiments


Internet, networks

Defining the EGEE Grid

- A Grid is the combination of networked resources and the corresponding middleware, which provides services for the user.

The EGEE Project

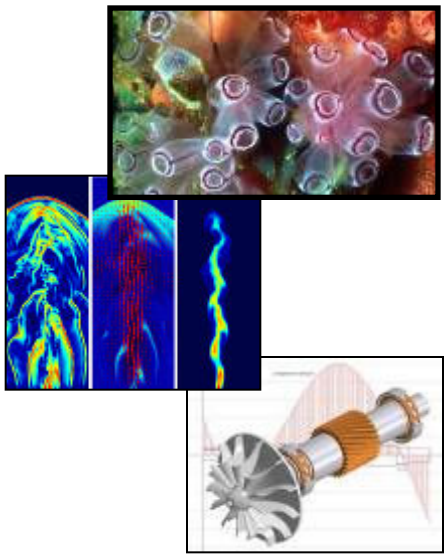
- **Aim of EGEE:**

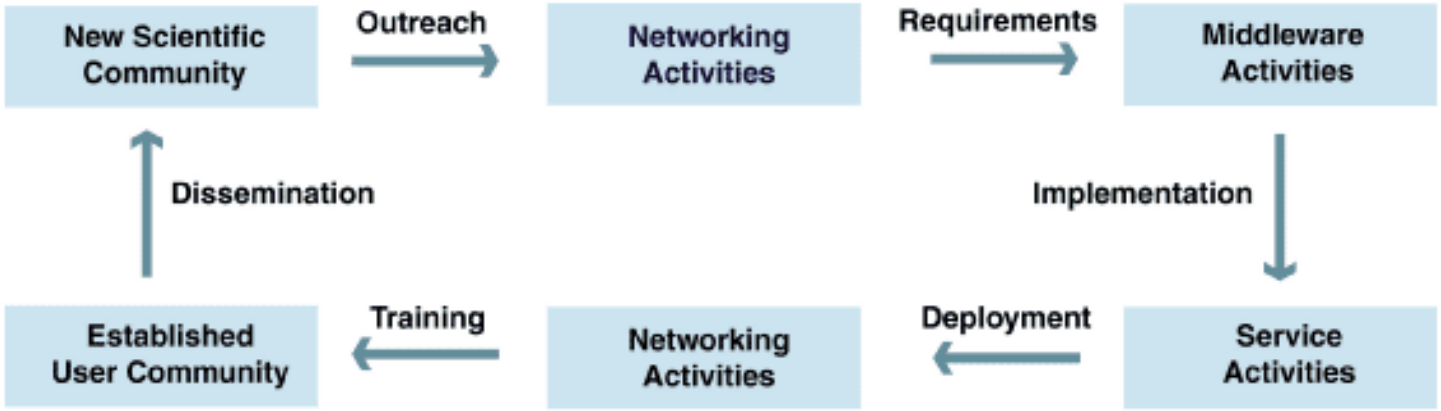
  "to establish a seamless European Grid infrastructure for the support of the European Research Area (ERA)"

- **EGEE**
  - 1 April 2004 – 31 March 2006
  - 71 partners in 27 countries, federated in regional Grids

- **EGEE-II**
  - 1 April 2006 – 30 April 2008
  - Expanded consortium

- **EGEE-III**
  - 1 May 2008 – 30 April 2010
  - Transition to sustainable model





Production infrastructure in numbers

17000 users

139,000 LCPUs (cores)

25Pb disk

39Pb tape

12 million jobs/month

+45% in a year

268 sites

+5% in a year

48 countries

+10% in a year

162 VOs

+29% in a year

Real Time Monitor:
http://gridportal.hep.ph.ic.ac.uk/rtm

Imperial College London

GridPP
UK Computing for Particle Physics

An EGEE site

- Computing service
- Storage service
- Security services

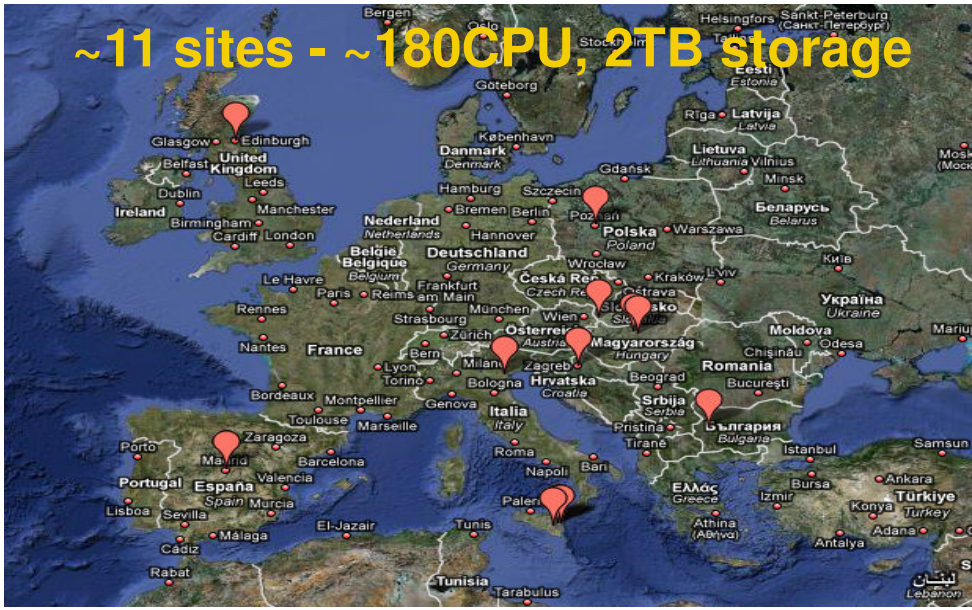
- grid.ucy.ac.cy
- IBM eServer 326m Machines
  - 2.0GHz dual AMD Opteron
  - 64-bit CPUs
  - 1GB RAM
  - 80 CPUs

EGEE Infrastructures

- **Production service**
  - Scaling up the infrastructure with resource centres around the globe
  - Running only well-tested and reliable middleware
  - 268 sites, 48 countries, ~139.000 CPUs (cores), ~25 PB storage
  - Separated into ~160 Virtual Organizations

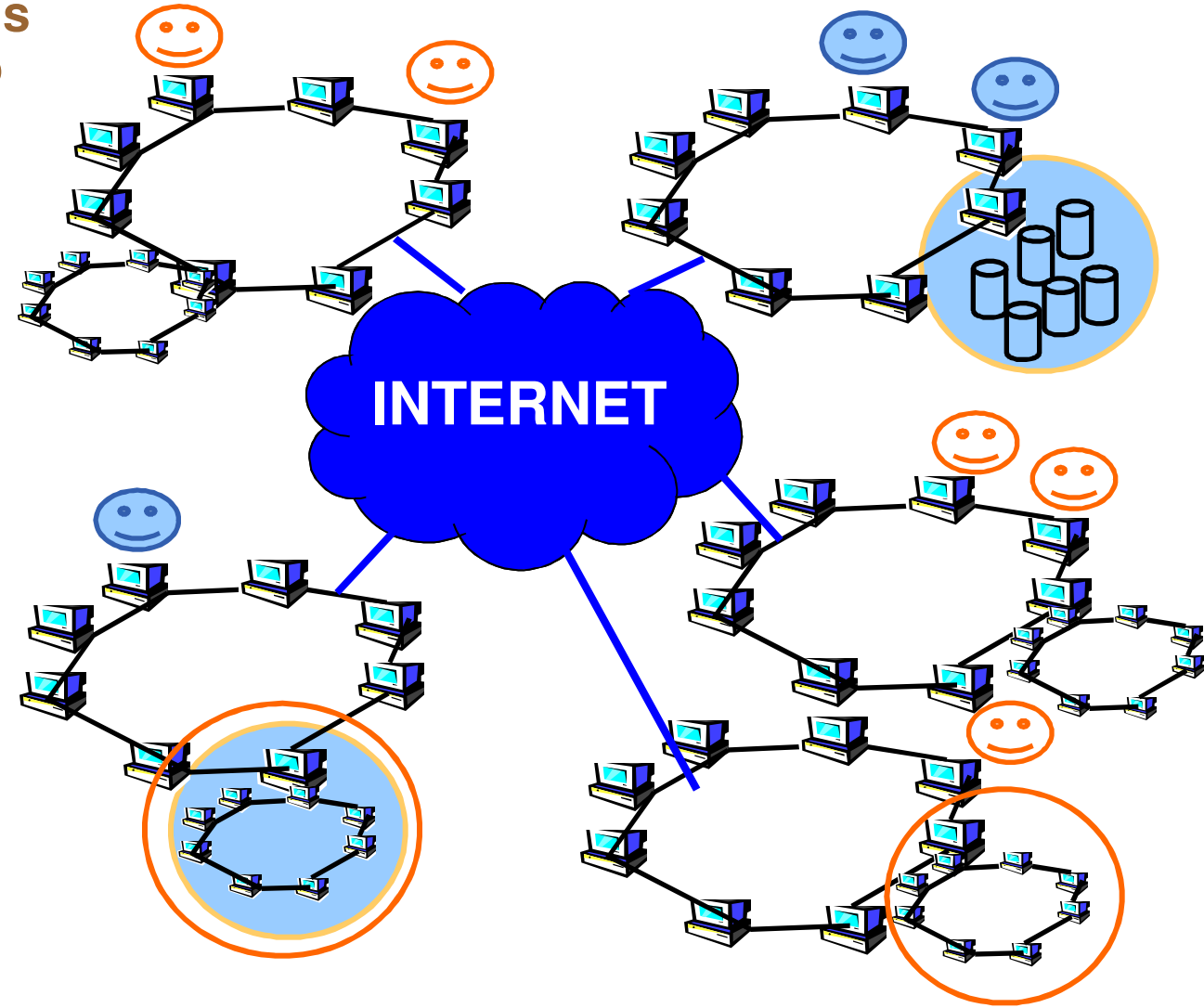
- **Pre-production service**
  - Run in parallel with the production service (restricted number of sites)
  - First deployment of new versions of the gLite middleware

- **Training-Infrastructure - GILDA**
  - Complete suite of Grid elements
  - Everyone can register and use GILDA for training and testing



EGEE, middleware and VOs

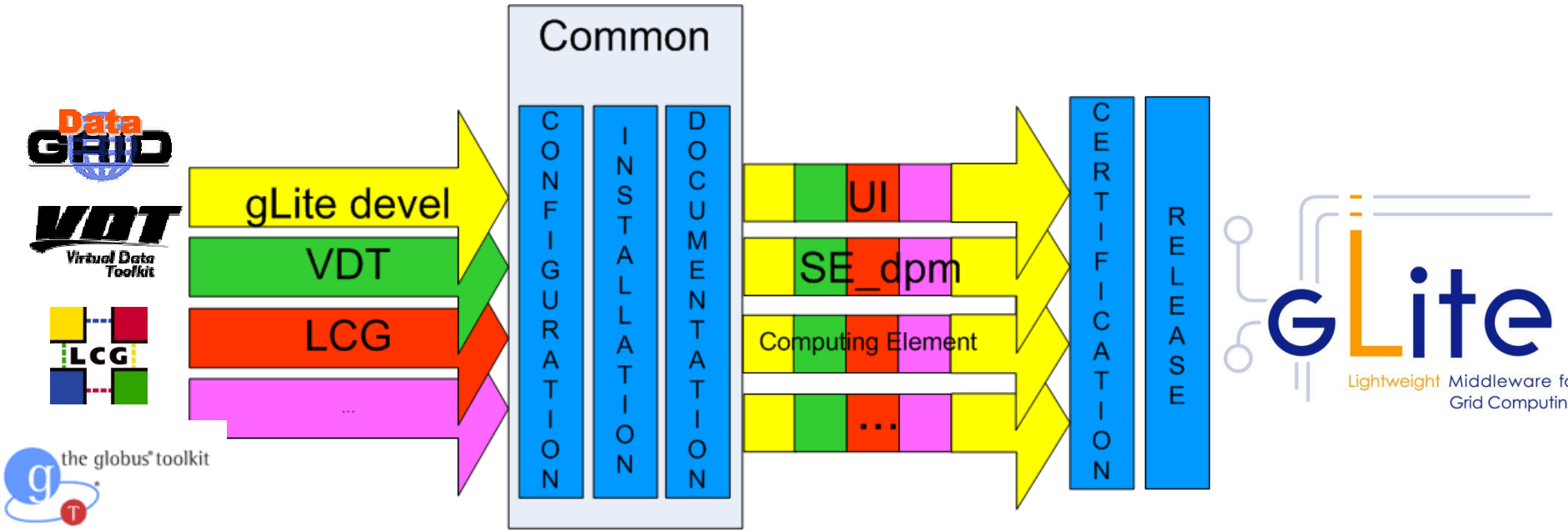
- **gLite middleware runs on each EGEE site to provide**
  - Data services
  - Computation services
  - Security service
- **EGEE resources and users form Virtual organisations: basis for collaboration**
- **Distributed services (both people and middleware) enable the grid**



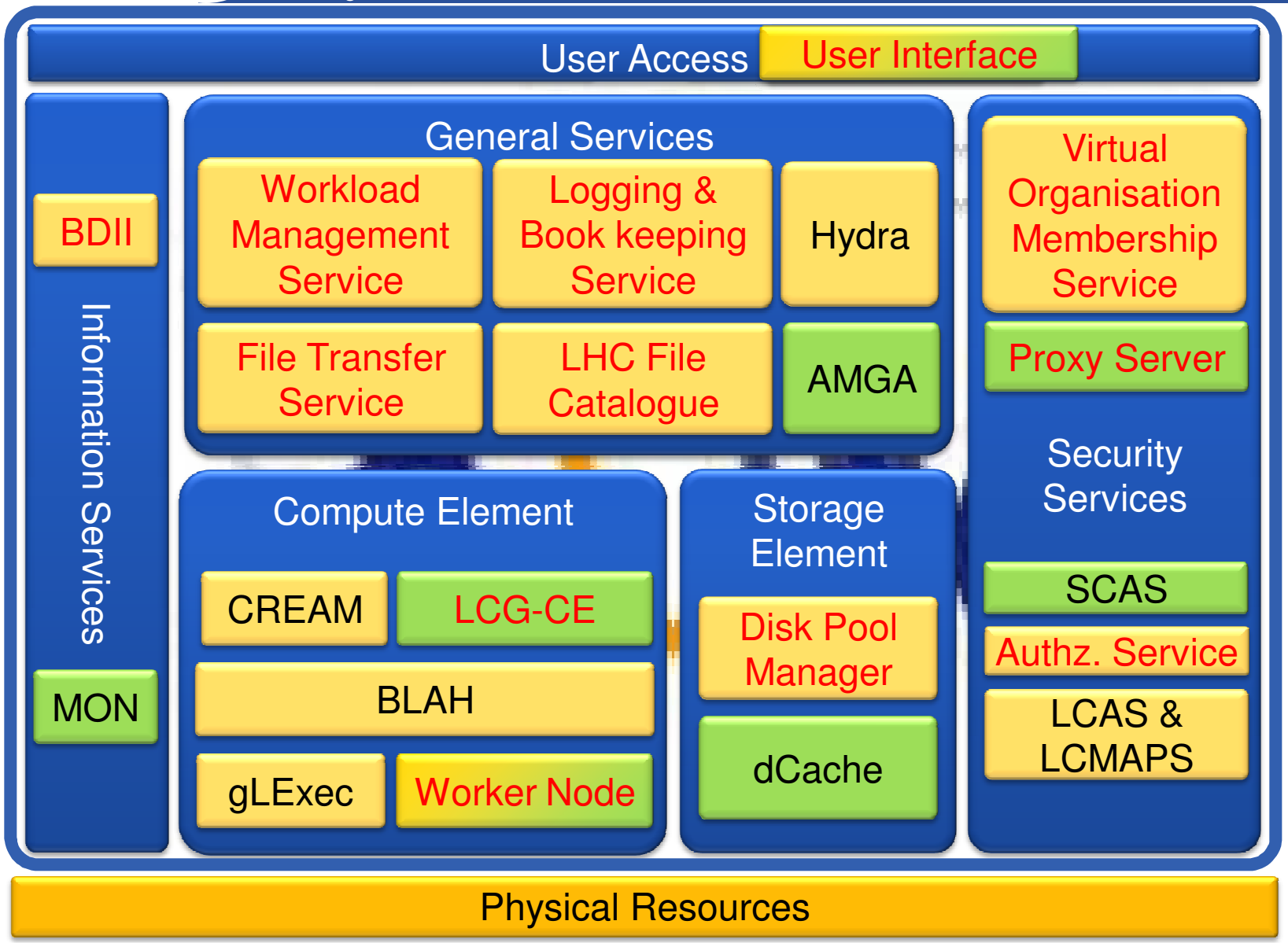
gLite: the EGEE middleware

www.glite.org

- **The current release is gLite 3.1**
- **gLite is…**
  - A distribution
  - With its own eponymous development project
- **It is updated almost every week**
- **Its purpose is to provide a stable platform for production grid usage**



The complete gLite service stack



Example: *Biomed* Virtual Organization

https://twiki.cern.ch/twiki/bin/view/EGEE/LifeSciences

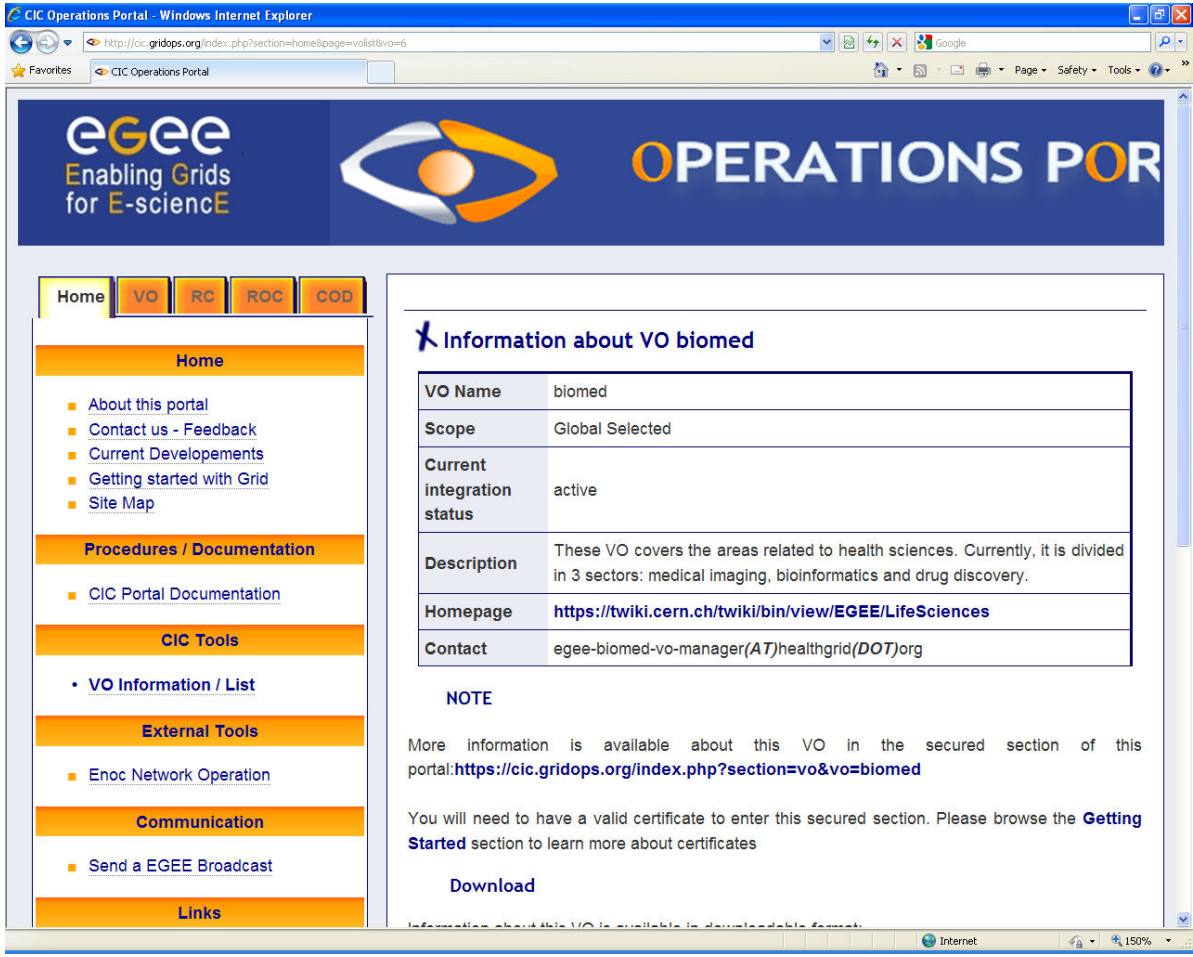
Sites:

- 36, mostly from Europe and Asia
- 65584 CPUs
- ~741 TB storage

Portal services are also available

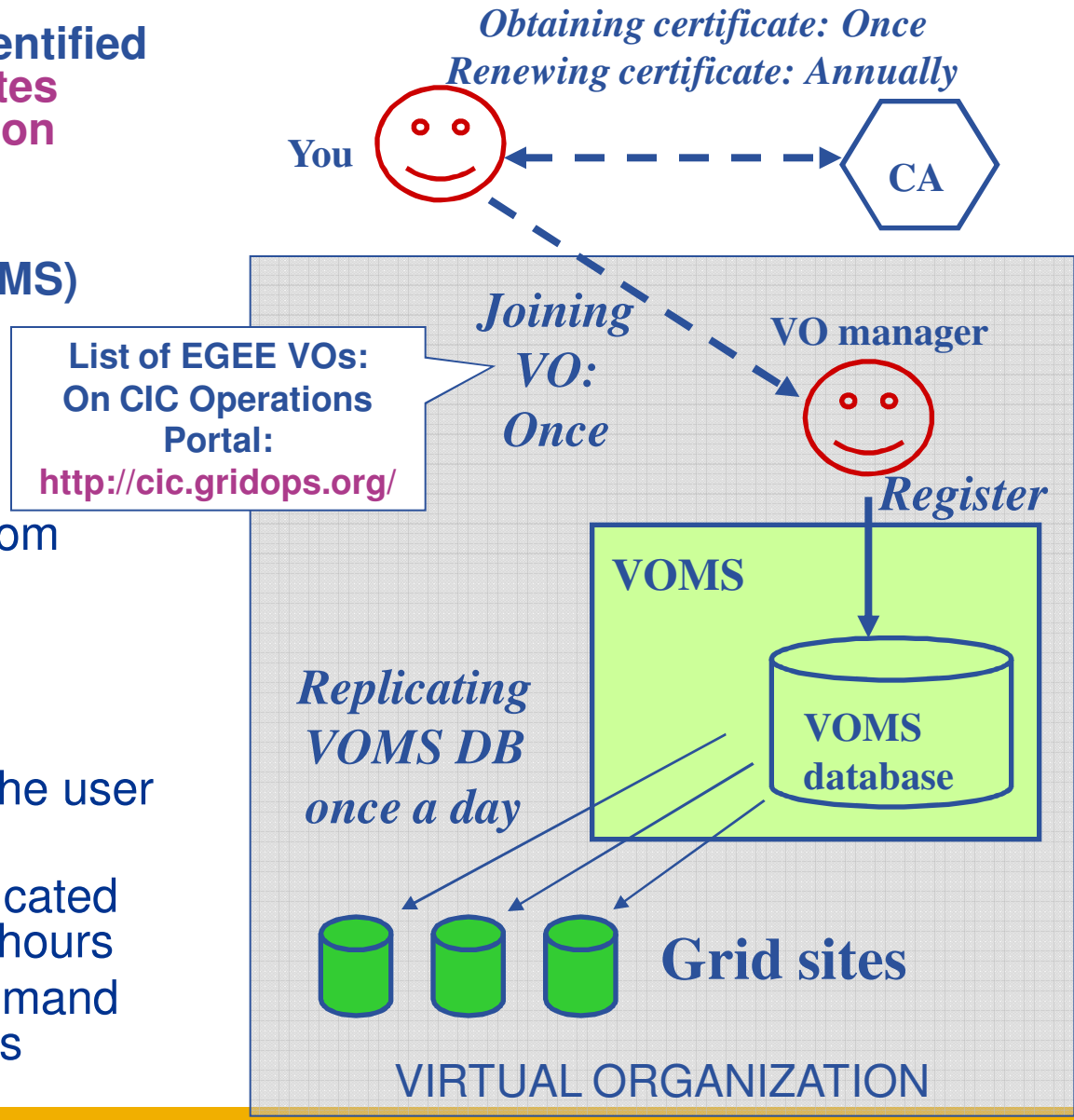
Members:

- Anybody working with biomedical applications
- No geographical restriction!
- Currently 216 persons



Getting access to an EGEE VO

- **Users (and machines) are identified by grid certificates. Certificates issued by national Certification Authorities (CA)**
- **VO Membership Service (VOMS) keeps list of people who are registered to a VO**
- **Users' steplist**
  - User obtains certificate from Certification Authority http://www.igtf.net
  - User registers at the VO
    - usually via a web form
  - VO manager authorizes the user
    - VOMS DB updated
  - The user's identity is replicated onto resources within 24 hours
  - Use the grid through command line or graphical interfaces



Find your national CA!

http://igtf.net

The International Grid Trust Federation

The international community is deploying large scale distributed computing grids on a production scale, across organisations, across countries, and across continents, for the advancement of science and engineering. In shaping this common grid infrastructure, many of these grids are relying on common practices, policies and procedures to reliably identify grid subscribers and resources.

News

The latest IGTF trust anchor distribution is always available from the PMA web sites: EUGridPMA and APGridPMA. Please refer to the README and CHANGES files for information about the distribution and its use.

The International Grid Trust Federation (IGTF) is a body to establish common policies and guidelines between its Policy Management Authorities (PMAs) members and to ensure compliance to this Federation Document amongst the participating PMAs. The IGTF does not provide identity assertions but instead ensures that within the scope of the IGTF charter the assertions issued by accredited authorities of any of its member PMAs meet or exceed an authentication profile relevant to the accredited authority.

Functions and the Trust Anchor Distribution

The IGTF maintains a list of trust anchors, root certificates and related meta-information for all the accredited authorities, i.e., those that meet or exceed the criteria mentioned in the *Authentication Profiles*. The Distribution contains Certificate Revocation List (CRL) locations, contact information, and signing policies.

- Download the latest update of the Common Distribution
- Download the Distribution Tools and the *fetch-crl* utility

Constituency

The IGTF constituency consists of our three member PMAs: the APGridPMA covering Asia and the Pacific, the EUGridPMA covering Europe, the Middle East and Africa, and The Americas Grid PMA covering Latin America, the Carribean and North America. All registered members in each regional PMA are also members of the IGTF. These include identity providers, CAs, and their major Relying Parties, such as the international Grid Deployment and Infrastructure projects.

About the IGTF

IGTF Charter
Tokyo Accord (2003)

Member PMAs and Registries

APGridPMA
EUGridPMA
TAGPMA
TACAR

Authentication Profiles

Classic X.509 CAs
Short-Lived Credential Services (SLCS)
Member Integrated Credential Services

Download the Distribution
Download the Utilities

Open Grid Forum Relationships

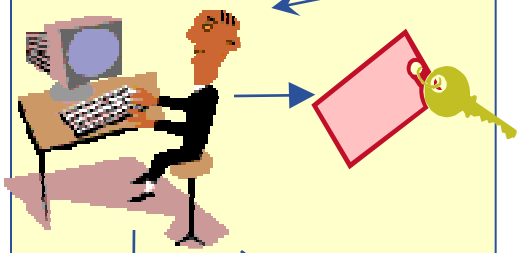
CA Operations WG
OGF Documents

Links

Open Grid Forum
TERENA TF-EMC2

Issuing a grid certificate

User generates public/private key pair in browser or in files.



Private Key encrypted on local disk:

passphrase

Cert Request Public Key

User sends public key to CA and shows proof of identity.

Certification Authority

CA signature links identity and public key in certificate.
CA informs user.

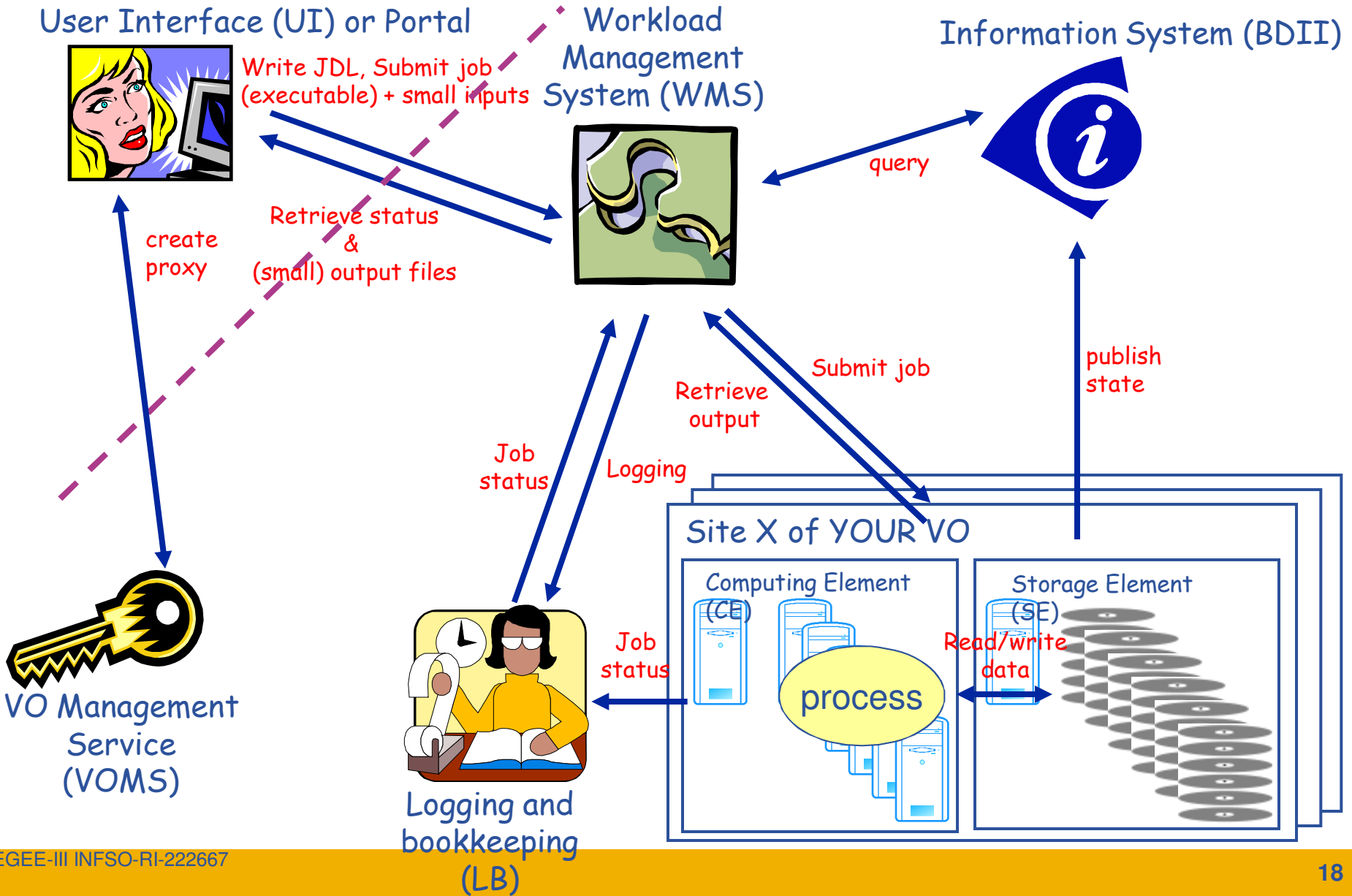
Cert

GILDA

https://gilda.ct.infn.it

- **We use the GILDA testbed for this event**
  - GILDA is a VO of EGEE → resources for training and application prototyping
  - A production EGEE VO looks like this!
  - Current EGEE production middleware: gLite 3.1.x
- **GILDA operates its own Certification Authority**
  - Training CA – lightweight authentication
  - No need for personal visit at GILDA CA!
  - GILDA renewable certificates are valid for 2 weeks (normal certificates for 1 year, also renewable)
- **Resources:**
  - ~12 sites
  - Central gLite services: VOMS, WMS, MyProxy, BDII, LB, ...
  - Portal interfaces:
    - P-GRADE Portal
    - GENIUS Portal
  - Training exercises: https://grid.ct.infn.it/twiki/bin/view/GILDA/WebHome
- **Ticketing system: http://gilda-support.ct.infn.it/**
  - Report problems about the infrastructure
  - Requires a separate registration
  - Login name / password based

Example usage scenario: Executing a single job



Main components

User Interface (UI): The place where users logon to the Grid

Resource Broker (RB) (Workload Management System (WMS): Matches the user requirements with the available resources on the Grid

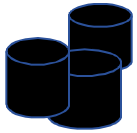
Int SE

Fil eplicas

Lo

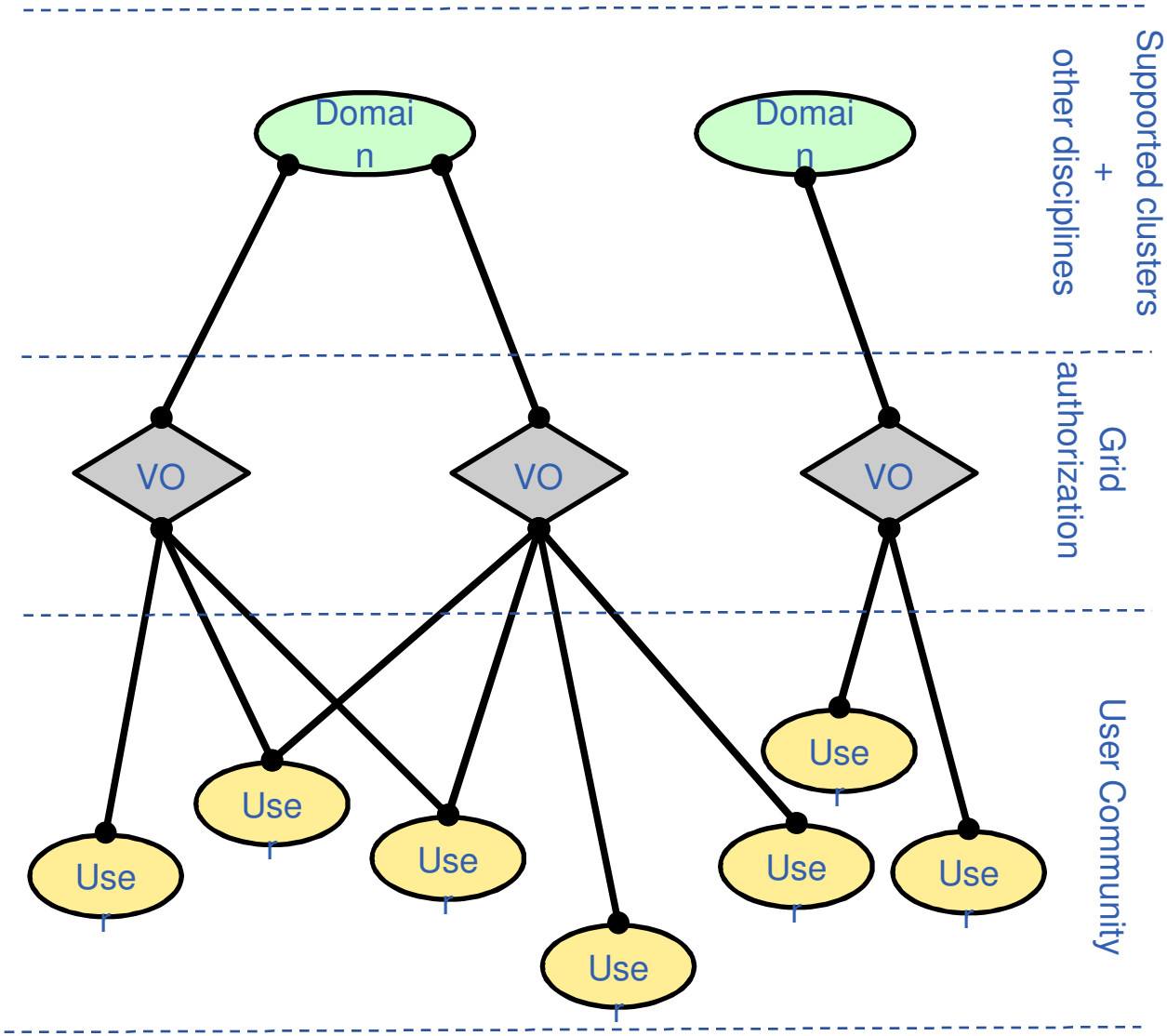
All built upon authorisation, authentication, security

Computing Element (CE): A batch queue on ere the user's job is executed



Storage Element (SE): provides (large-scale) storage for files

User Community Organization



Supported clusters:

- High Energy Physics (HEP)
- Life Sciences (LS)
- Earth Sciences (ES)
- Grid Observatory (GO, CS)
- Computational Chemistry (CC)
- Astronomy & Astrophysics (AA)
- Fusion (F)

23 "Core" VOs: CPU Utilization > 10% in the domain

112 total Registered VOs

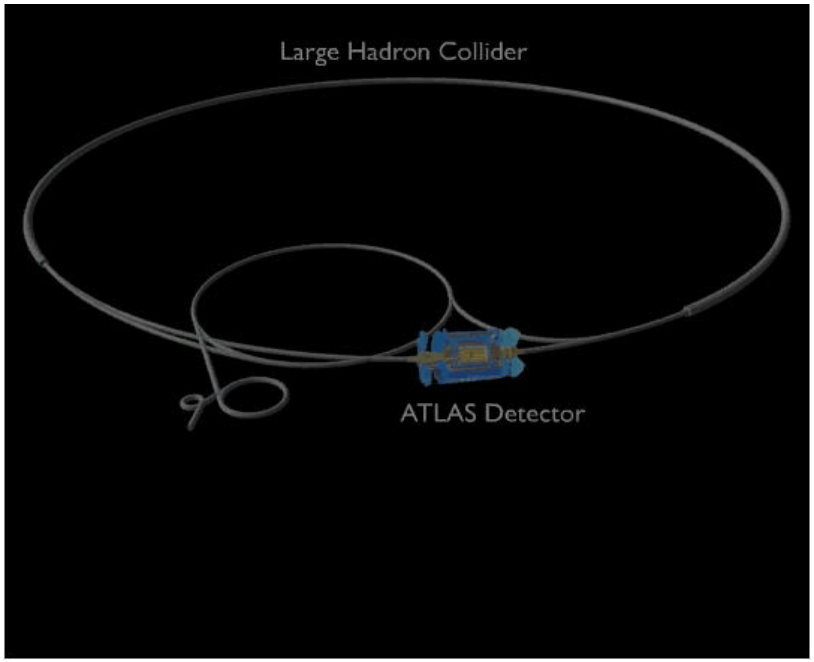
4167 users in "Core" VOs

13,381 users in Registered VOs

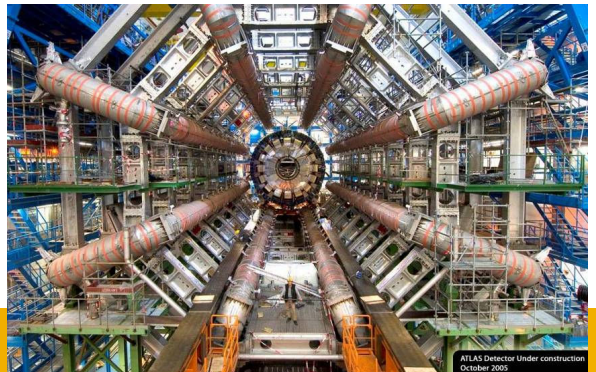
(Extreme) Application Example: LHC

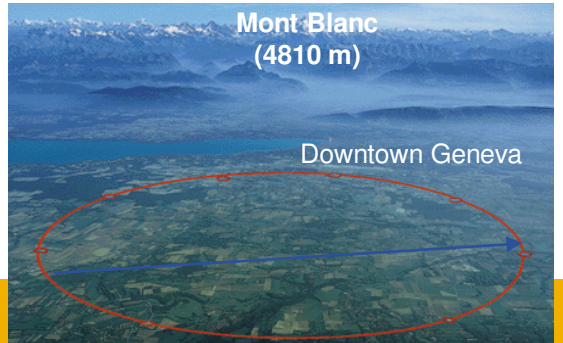
Brochure: http://cdsweb.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf

Large Hadron Collider

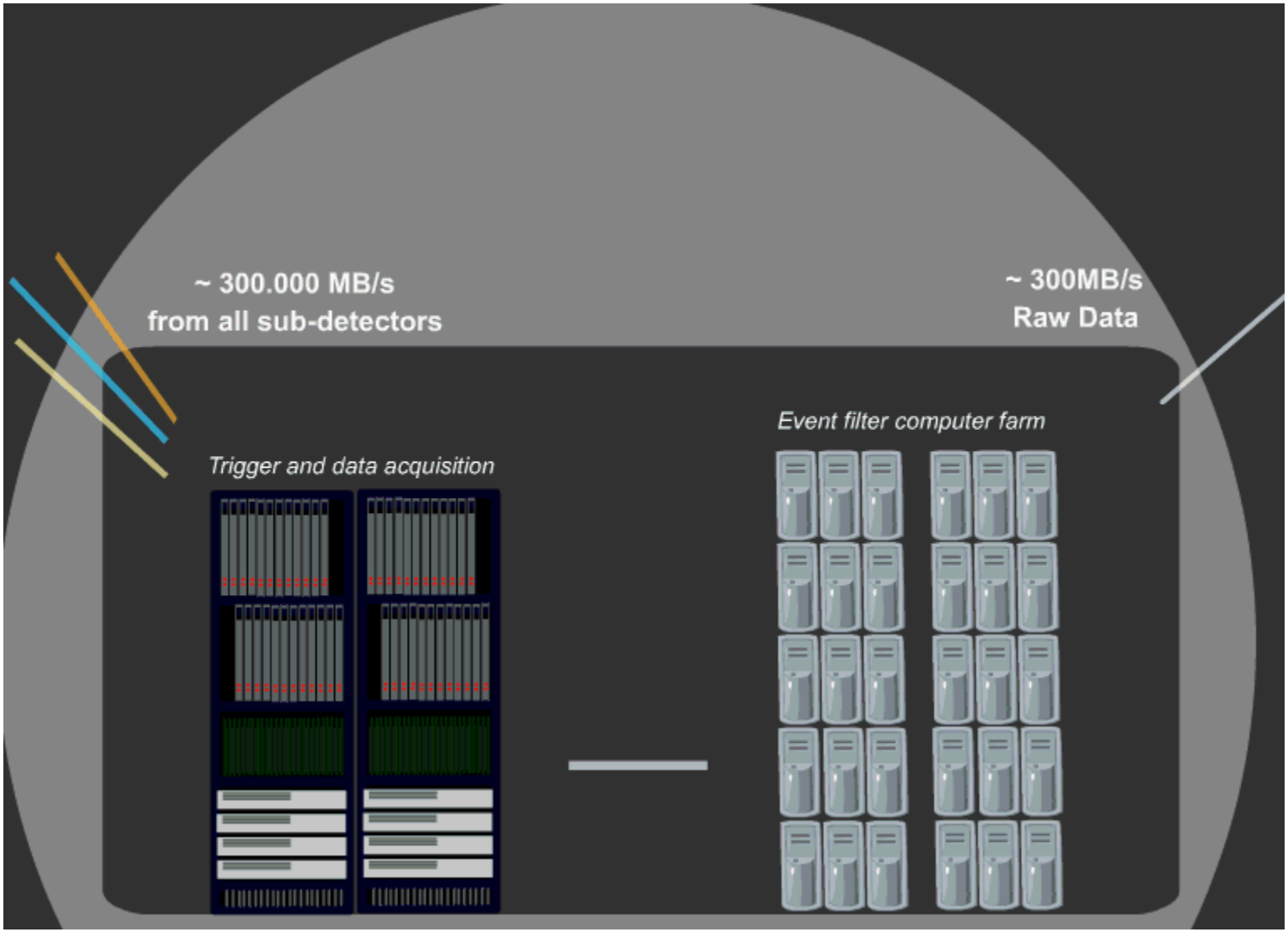


- **27 km circumference tunnel**
- **Initial startup in 2008, restart in 2009**
- **40 Million Particle collisions per second**
  - Online filter reduces to a few 100 “good” events per second recorded on disk and magnetic tape at 100-1,000 MegaBytes/sec
  - ~15 PetaBytes per year for all four experiments
- **Data analyzed by 100s of research groups world wide**

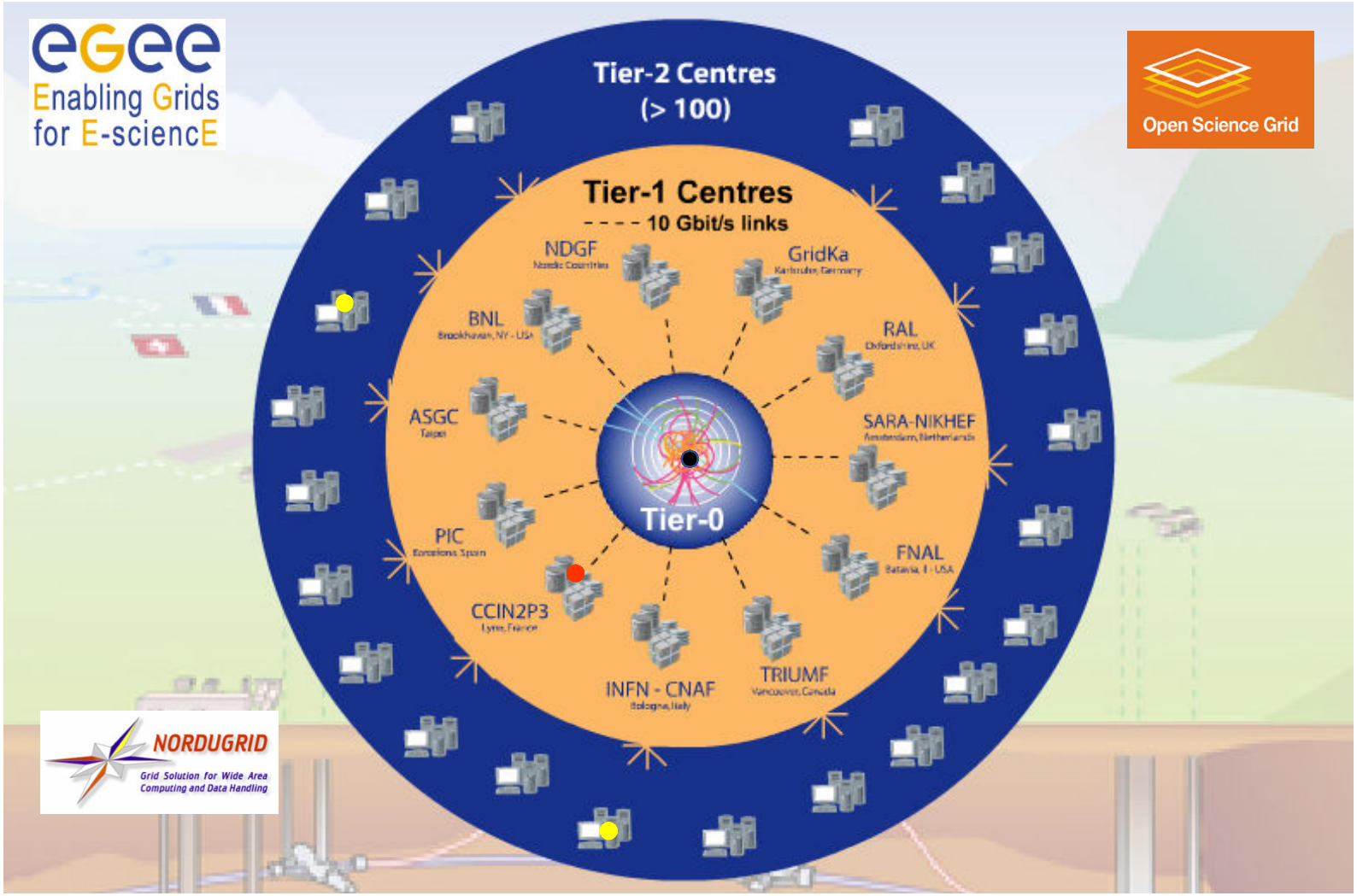




The Data Acquisition

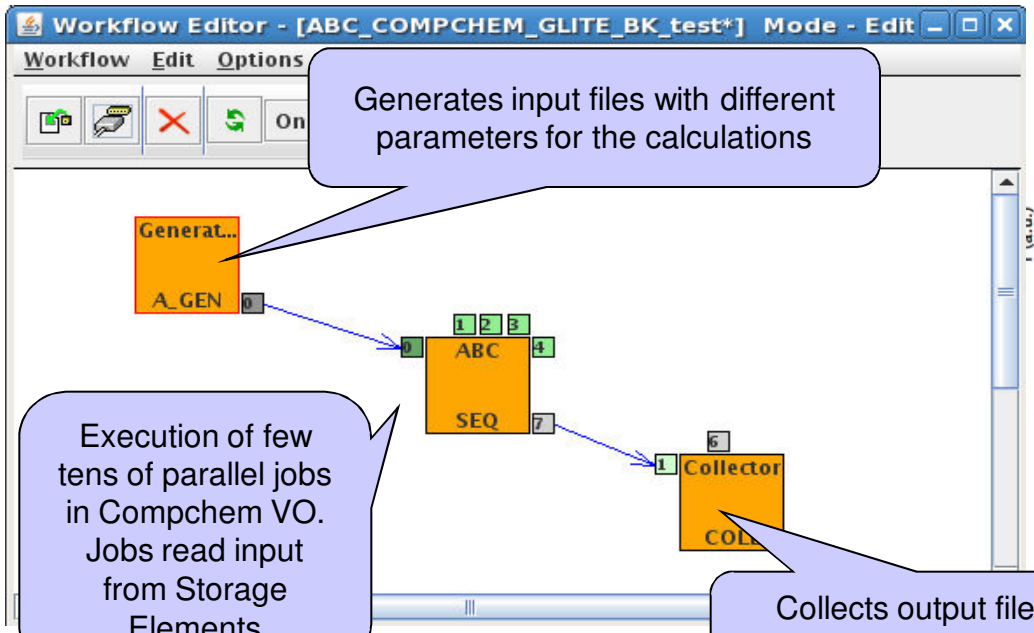


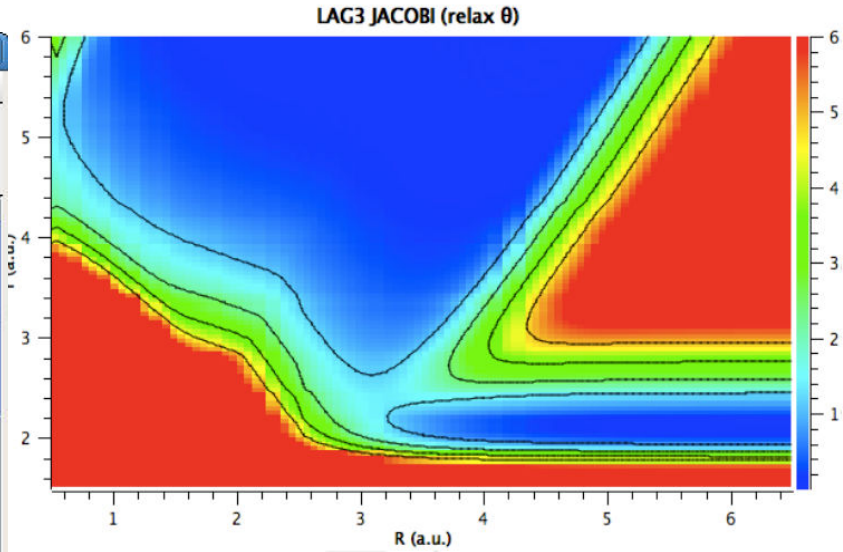
Data Distribution on the LHC-Grid



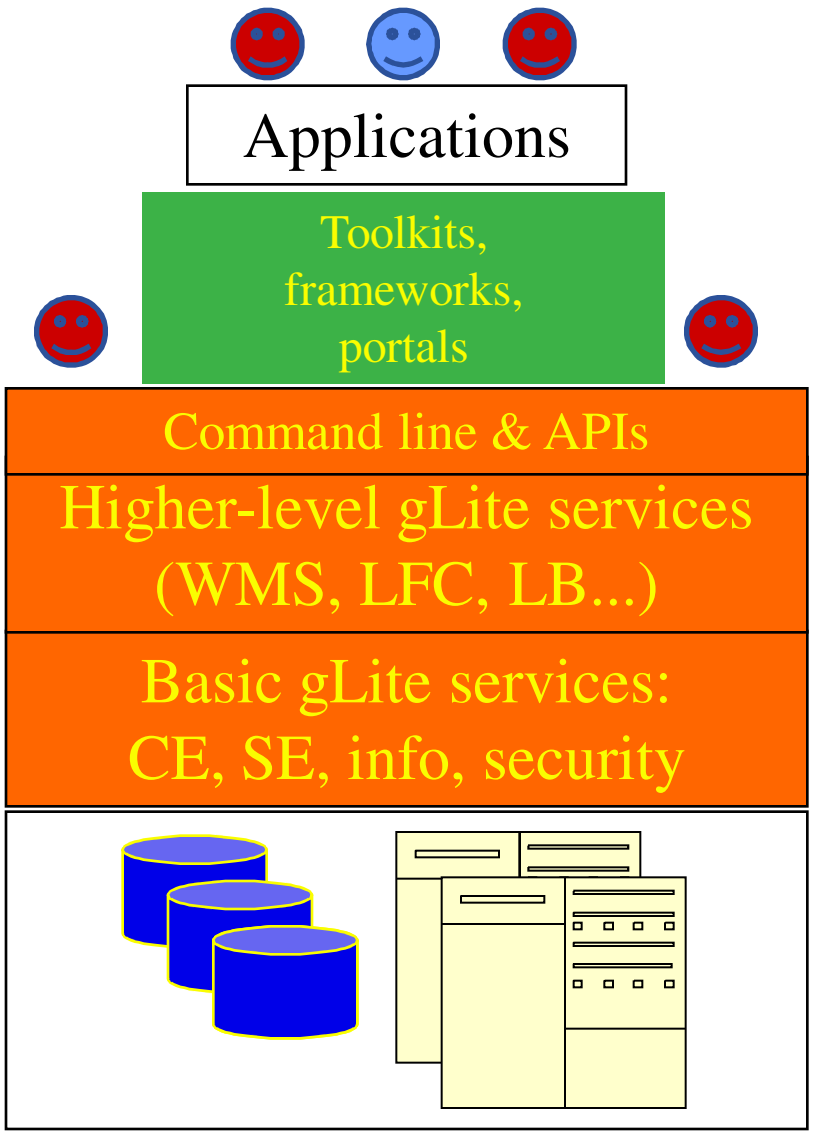
Compute grid application example: Computational Chemistry

- Solving the Schrodinger equation for triatomic systems
- Workflow of jobs
- Ported to EGEE with P-GRADE Portal
- Uses: WMS, CE, SE, LFC, BDII, VOMS





High level tools



- **Most VOs contain additional software services on top of gLite**
- **Some domain specific, some not**
- **EGEE RESPECT Programme:**
  - Recommended External Software Packages for Egee CommuniTies
  - Identify third-party software that works well with gLite
    - Expand the functionality of the grid infrastructure for users,
    - Reduce duplicated development when porting applications, and
    - Speed the porting of new application to the grid
  - http://technical.eu-egee.org/index.php?id=290

RESPECT

- **Identify third-party software that works well with gLite.**
  - http://technical.eu-egee.org/index.php?id=290

- **Simplified Access**
  - P-GRADE, Ganga, Migrating Desktop, g-Eclipse, i2glogin, Virtual Control Room
- **Workload Management**
  - GridWay Metascheduler, DIANE
- **New Resources**
  - GRelC, Instrument Element
- **Infrastructure Services**
  - StoRM

User Support Services

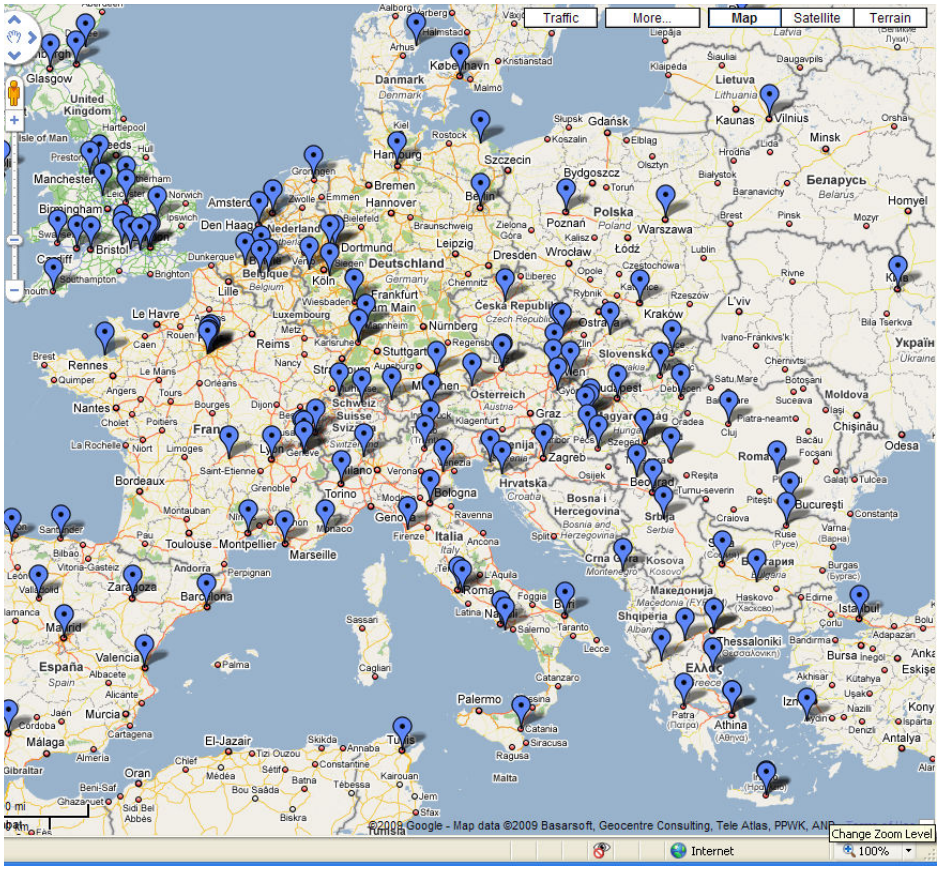
NA3 activity in EGEE: User training and induction

http://training.eu-egee.org/index.php?id=227

- Expand and refine the training material
- Train a wide variety of users
- Develop mechanisms for passing on knowledge to end-users of EGEE related projects
- Expand online learning services
- Support the provision of t-Infrastructure on the GILDA testbed

- Training material repositories:
  - http://egee.lib.ed.ac.uk/
    - Slides, Videos, MP3 talks on grid computing in general and EGEE
  - https://grid.ct.infn.it/twiki/bin/view/GILDA/UserTutorials
    - Tutorials for GILDA users

- Forthcoming and past training events
  - http://www.egee.nesc.ac.uk/



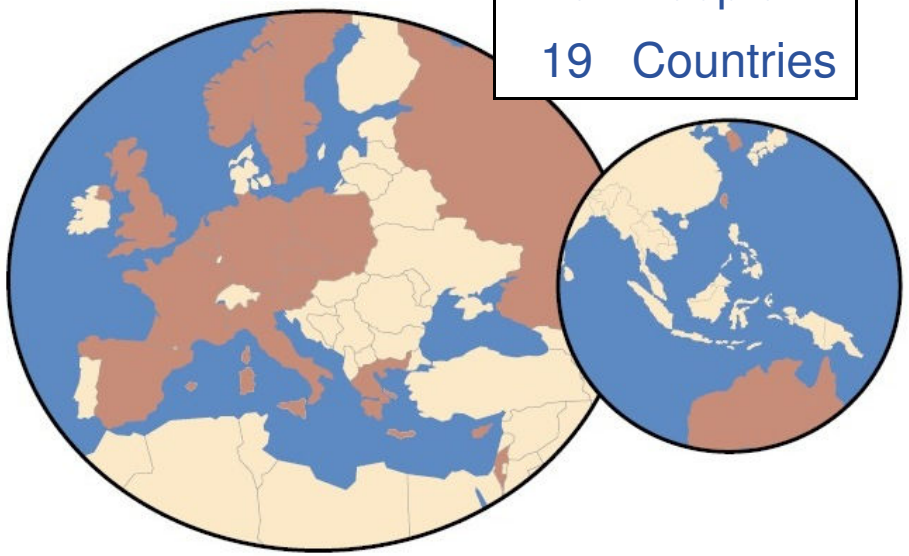
NA4 Activity in EGEE: User community expansion and support

- **Support the large and diverse EGEE user community:**
  - **Promote dialog**: Users' Forums & EGEE Conferences
  - **Technical Aid**: Porting support, procedural issues
  - **Liaison**: Software and operational requirements

- **Main support services: (MORE LATER!)**
  - Virtual Organization Support
  - Application Porting Support
  - Direct User Support

- **Strategic Discipline Clusters:**
  - High Energy Physics
  - Life Sciences
  - Earth Sciences
  - Grid Observatory
  - Computational Chemistry
  - Astronomy & Astrophysics
  - Fusion

44 Partners
28 People
19 Countries



Application Database

http://appdb.eu-egee.org

EGEE – RS Applications Registry

CIC portal EGEE

Mac ▾ Entertainment ▾ Tools ▾ Workshops ▾ AWS-Info ▾ Perl::Critic Bonnie Snapz

EGEE Applica

Home | User | Download | Admin

You are in: Region: EGEE >> User >> Registry

Records shown 1 - 10 of 220

Custom Filter | All Fields | Whole words only | Apply Filter | Reset Filter

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 Next >>

			Id	Logo	Acronym	Application Name	Discipline	Sub-Discipline	Description
View			48		LAWEF-GRID	Limited Area Weather and Environmental Forecasting on a Grid Platform	Earth Sciences	Meteorology	This applicat warnings rela events. Such approached weather fore periods (18 h the concept integrated or km) covering Mediterranea computationa fashion unde

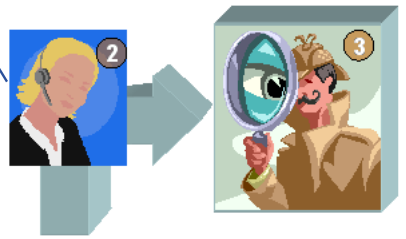
Id	87
Logo	
Acronym	ISDEP
Application Name	Integrator of Stochastic Differential Equations for Plasmas
Discipline	Fusion
Sub-Discipline	Statistical mechanics
Description	The purpose of this code is solving the collisional transport in fusion plasmas using the equivalence between the Fokker-Planck and Langevin equation (Stochastic Differential Equation). Altought in principle it solves the linearized kinetic equation, we include non linear terms using an iterative method. Several features can be taken into account in the equations, for example the presence of divertors and ion heating (ICH and NBI). This code has been succesfully used in the TJ-II stellarator (CIEMAT) and is currently working with ITER. As it is a Montecarlo sequencial code, it scales perfectly in clusters with slow connections between nodes. ISDEP has been ported to the EGEE grid and Ibercivis, a volunteer computing platform. It has also been used in int.eu.grid project as computing core for interactive visual fusion simulations (IVISDEP) and is taking part in other grid projects like EDGeS and EUFORIA. The main advantages of ISDEP is the absence of common approximations assumed in other codes. For example, there is no assumption in the complexity of the system geometry.
Scientific Contact	View ...
Web Site	
Region	SWE
VO	fusion

Alt. link: http://grid.ct.infn.it/egee_applications/

Application Porting Support

- **Consultancy service on how to port Your application to EGEE**
- **Coordinator: MTA SZTAKI, Budapest**
  - Grid Application Support Center (GASuC)
    - http://www.lpds.sztaki.hu/gasuc
  - Leader:
    - Gergely Sipos (sipos@sztaki.hu)
    - Budapest, XIII. Victor Hugo utca 18-22.
- **Other porting centres:**
  - INFN Catania
  - CSIC Santander
  - UCM Madrid (GridWay)
  - ASGC Taipei
  - UoM Melbourne
- **Other partners**
  - WMIN London: http://wgrass.wmin.ac.uk/index.php/W-Grass_team

Support cycle and services



Problem analysis

Interviews

Personalized training

Apply online at www.lpds.sztaki.hu/gasuc

Porting specifications

Writing publications and case studies

Prototyping grid applications

Fine tuning applications on production grids

Visit porting centre to make the cycle more effective!

EGEE Community Building Fund can subsidize your travel

Operational Support: GGUS

- **Operational support deals with day-to-day problems of using the EGEE infrastructure.**
- **GGUS (Global Grid User Support)**
  - Ticketing handling system with many specialized teams.
  - URL: http://ggus.org/
  - **Grid certificate is needed to access the site!**
- **Speeding response time:**
  - Investigate as much as possible yourself.
  - Try to identify service or site causing problems.
  - Provide as much information as possible.
  - Complain…

- **Even though ticketing systems have many benefits, it is often more efficient to have one-on-one interaction with a knowledgeable person.**
  - Local site administrator:
    - Usually very knowledgeable about grid.
    - Very motivated to ensure that local users are happy.
  - NA4 partners:
    - Many partners distributed throughout Europe.

Summary

Grids represent a powerful new tool for science

→Today we have a **window of opportunity** to move grids from research prototypes to permanent production systems (as networks did a few years ago)

EGEE offers …

- … a mechanism for **linking together people**, **resources and data** of many scientific community
- … a basic set of middleware for *gridfying* applications with documentation, training and support
- … regular forums for linking with grid experts, other communities and industry

Not only for large communities!
You can also use this technology!

Further information

- EGEE
  - http://www.eu-egee.org/
- gLite middleware
  - http://www.glite.org
- gLite manuals, documentation
  - http://glite.web.cern.ch/glite/documentation/ (gLite user guide)
- Recommended External Software Packages for Egee CommuniTies
  - http://technical.eu-egee.org/index.php?id=290

Thank you