



Enabling Grids for E-science

Overview of the EGEE project and middleware

Presented by Gergely Sipos, sipos@sztaki.hu

With thanks to EGEE colleagues for many of these slides

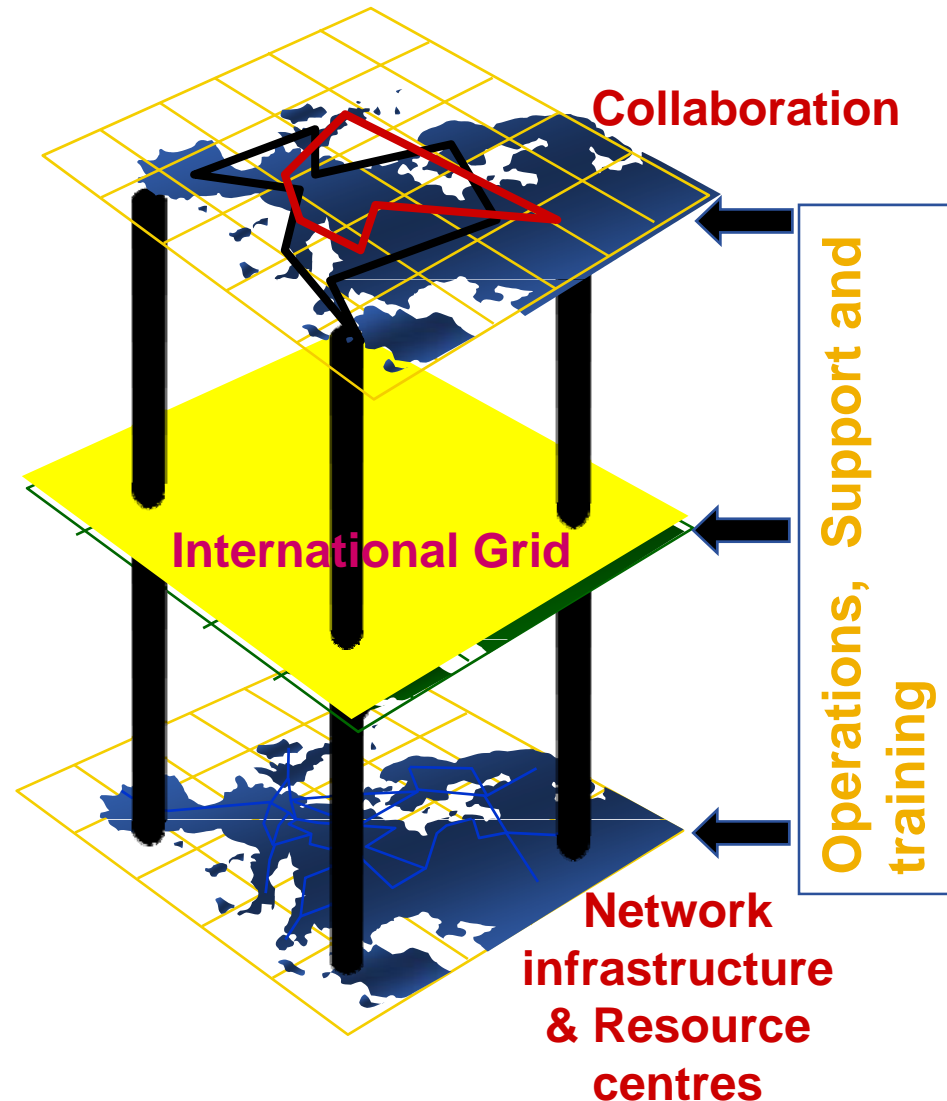
www.eu-egee.org



- **What is EGEE?**
 - Goals
 - Status
 - Activities
- **EGEE's Grid middleware: gLite 3**
- **Sources of further information**



- **Build, deploy and operate a consistent, robust a large scale production grid service that**
 - Links with and build on national, regional and international initiatives
- **Improve and maintain the middleware in order to deliver a reliable service to users**
- **Attract new users from research and industry and ensure training and support for them**



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

- **From April 2006, natural continuation of EGEE**

- Expanded consortium
- Emphasis on providing an infrastructure
 - increased support for applications
 - interoperate with other infrastructures
 - more involvement from Industry

SA: service activities

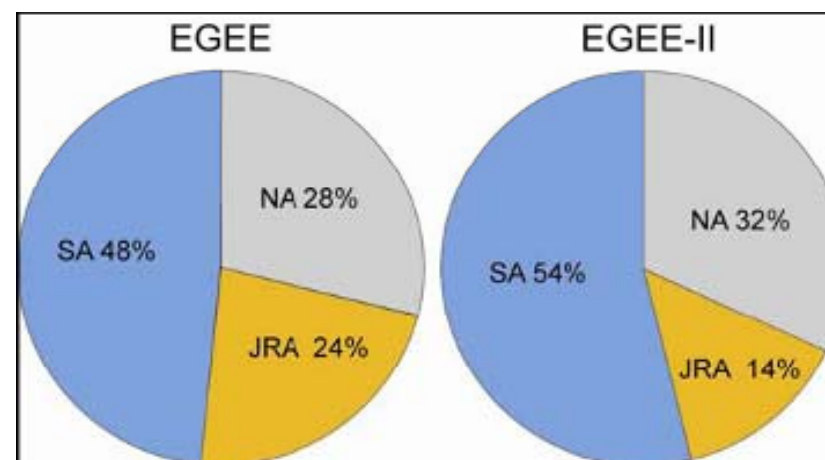
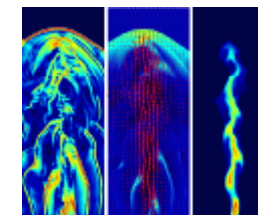
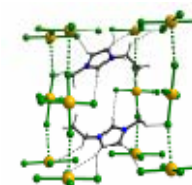
- establishing operations

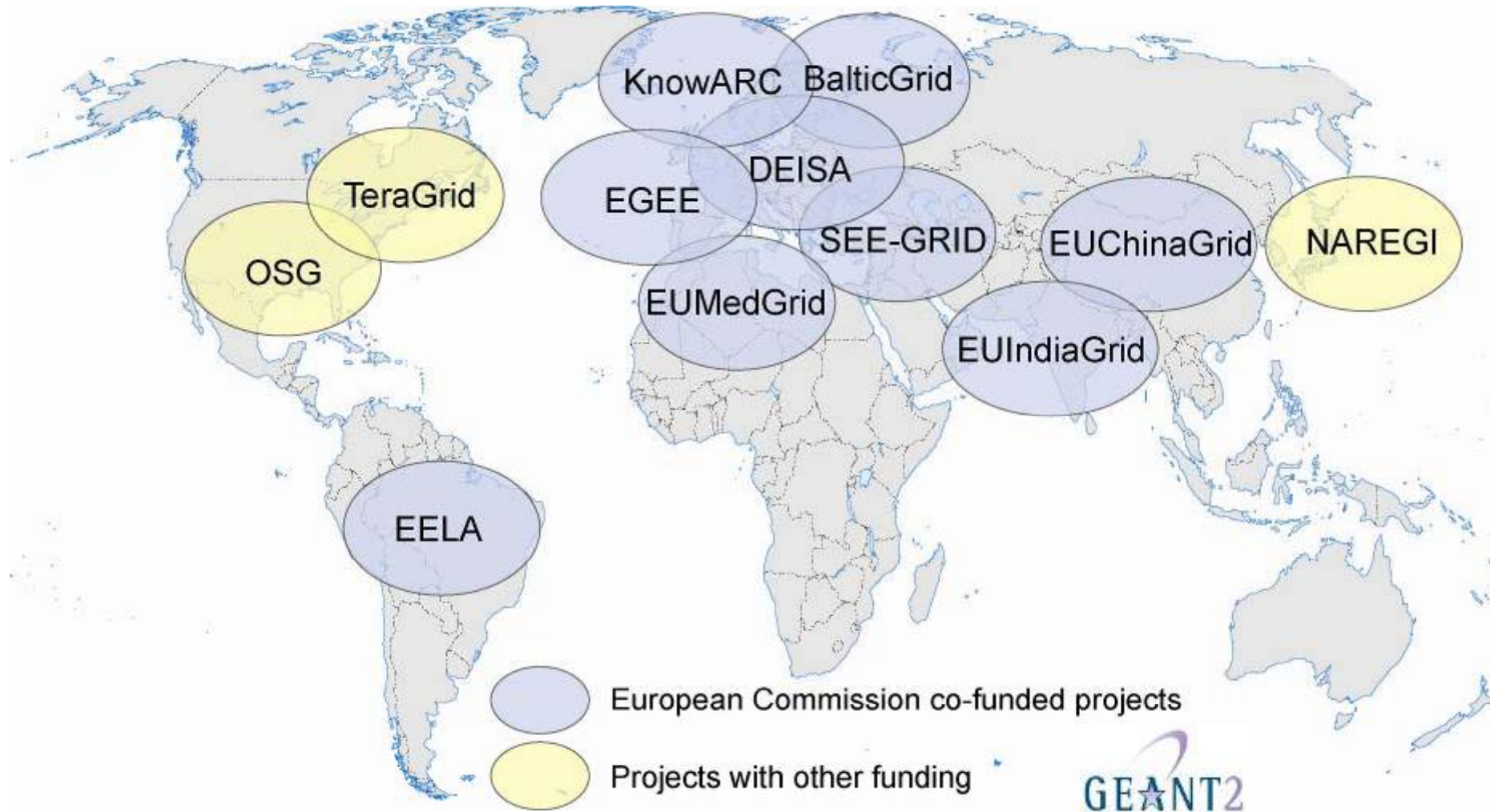
NA: network activities

- supporting VOs

JRA: “joint research activities”

- e.g. hardening middleware





Potential for linking ~80 countries by 2008

<i>Name</i>	<i>Description</i>
BalticGrid	EGEE extension to Estonia, Latvia, Lithuania
EELA	EGEE extension to Brazil, Chile, Cuba, Mexico, Argentina
EUChinaGRID	EGEE extension to China
EUMedGRID	EGEE extension to Malta, Algeria, Morocco, Egypt, Syria, Tunisia, Turkey
ISSeG	Site security
eIRGSP	Policies
ETICS	Repository, Testing
OMII-Europe	to provide key software components for building e-infrastructures;
BELIEF	Digital Library of Grid documentation, organisation of workshops, conferences
BIOINFOGRID	Biomedical
Health-e-Child	Biomedical – Integration of heterogeneous biomedical information for improved healthcare
ICEAGE	International Collaboration to Extend and Advance Grid Education

Test-beds & Services

Certification testbeds (SA3)

Pre-production service

Production service

Infrastructure:

- Physical test-beds & services
- Support organisations & procedures
- Policy groups

Support Structures

Operations Coordination Centre

Regional Operations Centres

Global Grid User Support

EGEE Network Operations Centre (SA2)

Operational Security Coordination Team

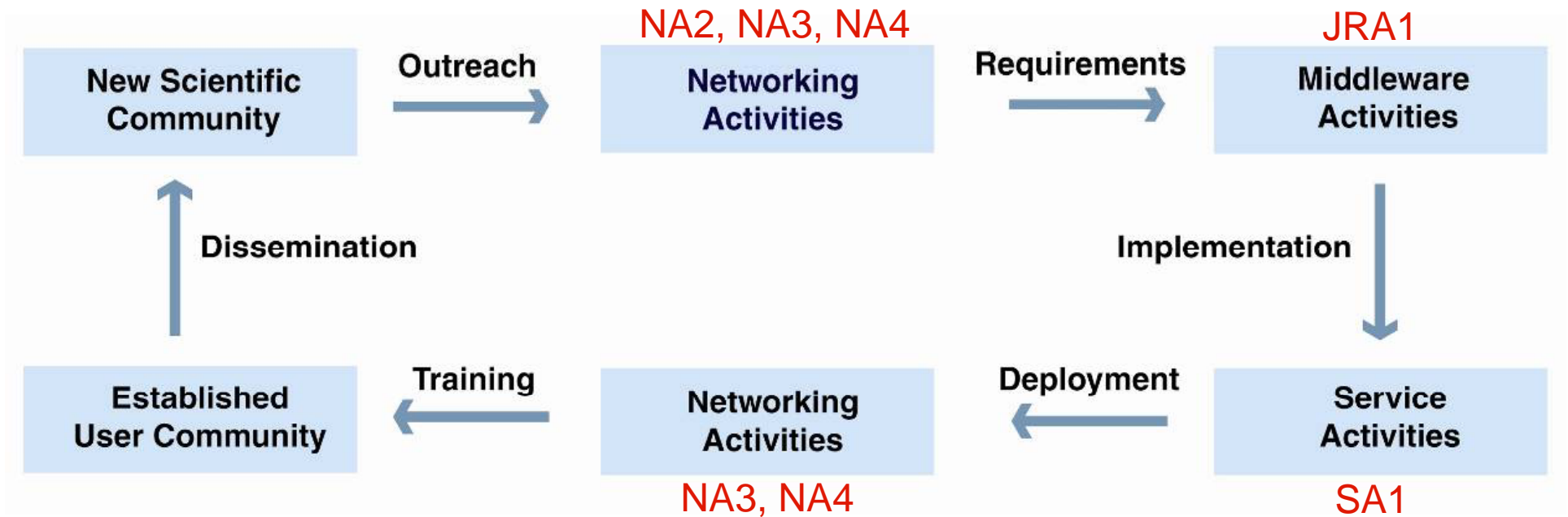
Security & Policy Groups

Joint Security Policy Group

EuGridPMA (& IGTF)

Grid Security Vulnerability Group

Operations Advisory Group (+NA4)



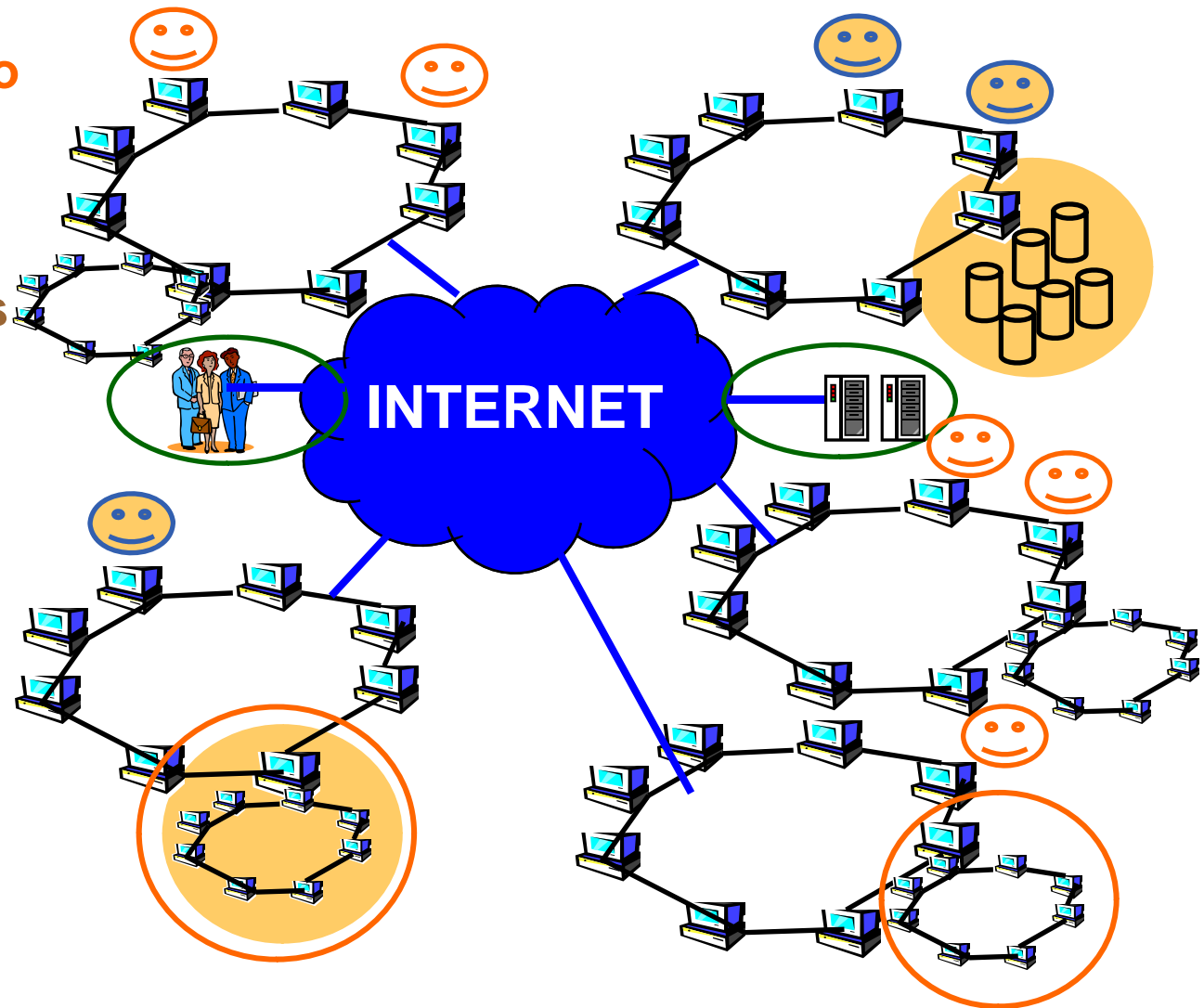
Building effective user communities

- **What is EGEE?**
 - Goals
 - Status
 - Activities
- **EGEE's Grid middleware: gLite 3**
- **Sources of further information**



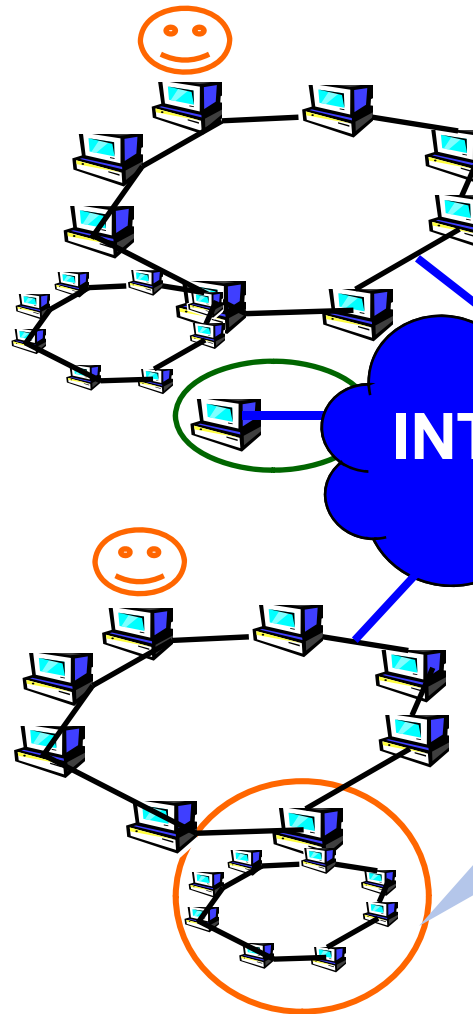
- **What is a Virtual Organisation?**
 - People in different organisations seeking to cooperate and share resources across their organisational boundaries
 - E.g. A research collaboration
- **Each grid is an infrastructure enabling one or more “virtual organisations” to share and access resources**
- **Each resource is exposed to the grid through an abstraction that masks heterogeneity, e.g.**
 - Multiple diverse computational platforms
 - Multiple data resources
- **Resources are usually owned by VO members. Negotiations lead to VOs sharing resources**

- **Virtual organisations negotiate with sites to agree access to resources**
- **Grid middleware runs on each shared resource to provide**
 - Data services
 - Computation services
 - Single sign-on
- **Distributed services (both people and middleware) enable the grid**



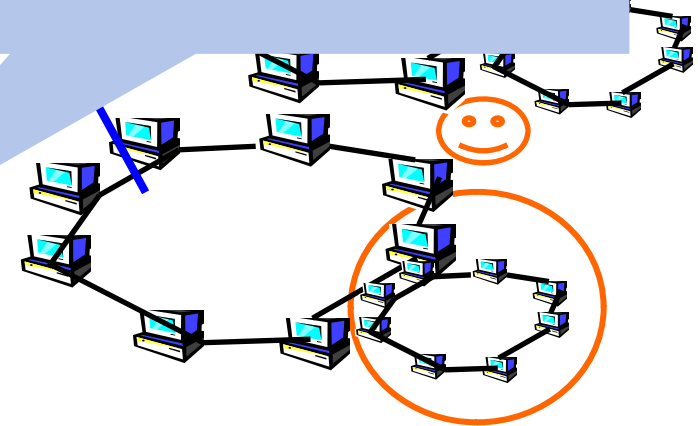
Typical current grid

- **Grid middleware runs on each shared resource**
 - Data storage
 - (Usually) batch queues on pools of processors
- **Users join VO's**
- **Virtual organisation negotiates with sites to agree access to resources**
- **Distributed services (both people and middleware) enable the grid, allow single sign-on**



At each site that provides computation:

- Local resource management system
- (= batch queue)
 - PBS
 - ...
- EGEE term: queue is a "Computing element"



- **When using a PC or workstation you**
 - Login with a username and password (“Authentication”)
 - Use rights given to you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **Operating system**
- **One admin. domain**
- **When using a Grid you**
 - Login with digital credentials – single sign-on (“Authentication”)
 - Use rights given you (“Authorisation”)
 - Run jobs
 - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**



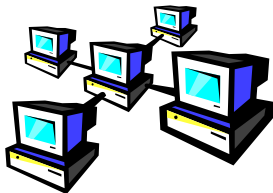
User Interface (UI): The place where users logon to the Grid



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



Information System: Characteristics and status of CE and SE



Computing Element (CE): A batch queue on a site's computers where the user's job is executed



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



User Interface (UI):

The place where users logon to the Grid



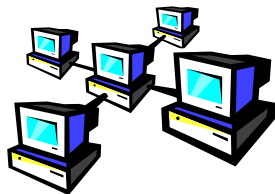
Resource Broker (RB): Matches the user requirements with the available

**All built upon
authorisation,
authentication,
security**



Int

SE



C

ere



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

Grid Security Infrastructure: Single logon with security and trust



User Interface (UI):

The place where users logon to the Grid



Resource Broker (RB):

Matches the user requirements with the available resources on the Grid



Information System:

Characteristics and status of CE and SE
(Uses “GLUE schema”)



Computing Element (CE):

A batch queue on a site’s computers where the user’s job is executed



Storage Element (SE):

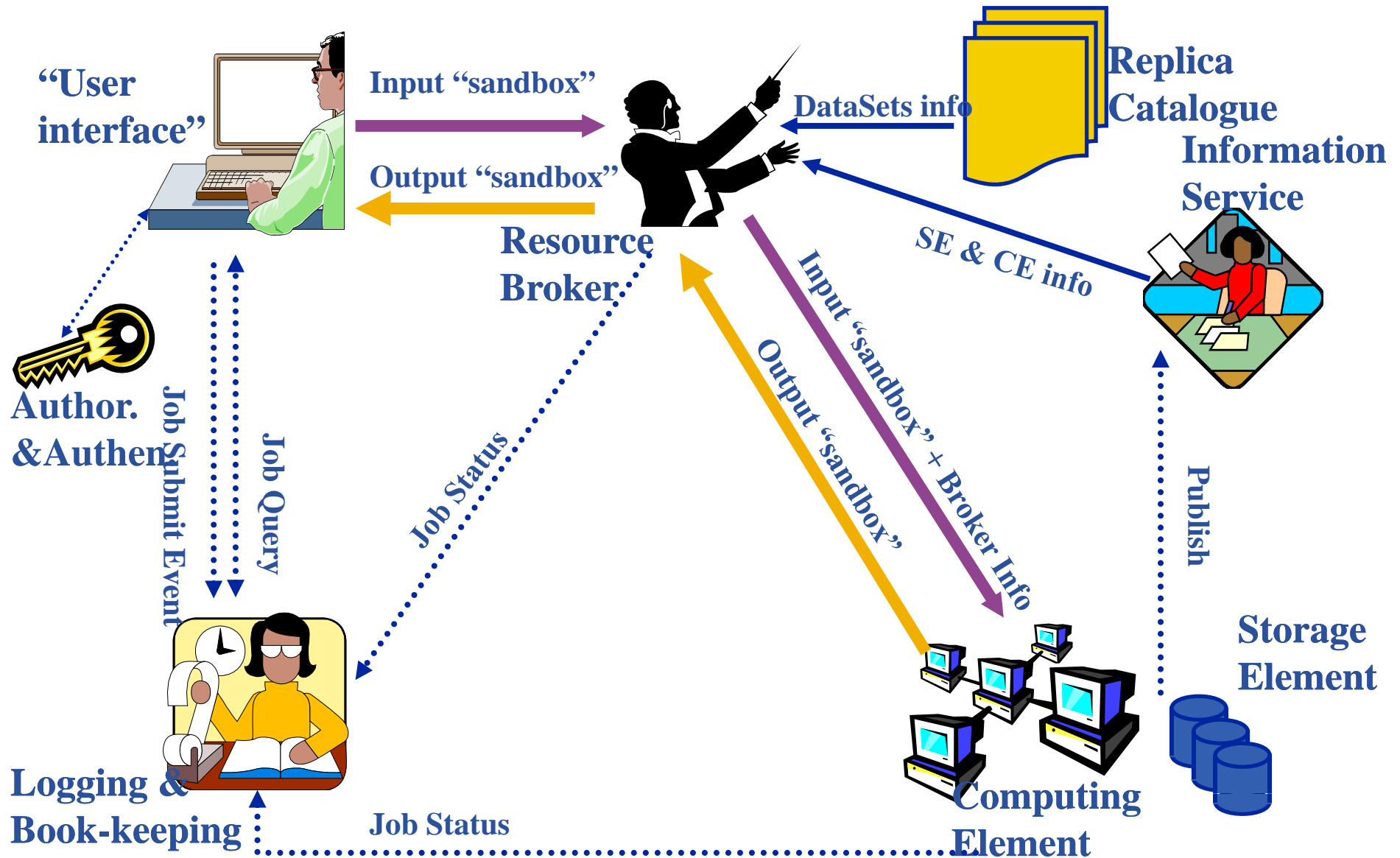
provides (large-scale) storage for files



Grid Security Infrastructure:

Single logon with security and trust

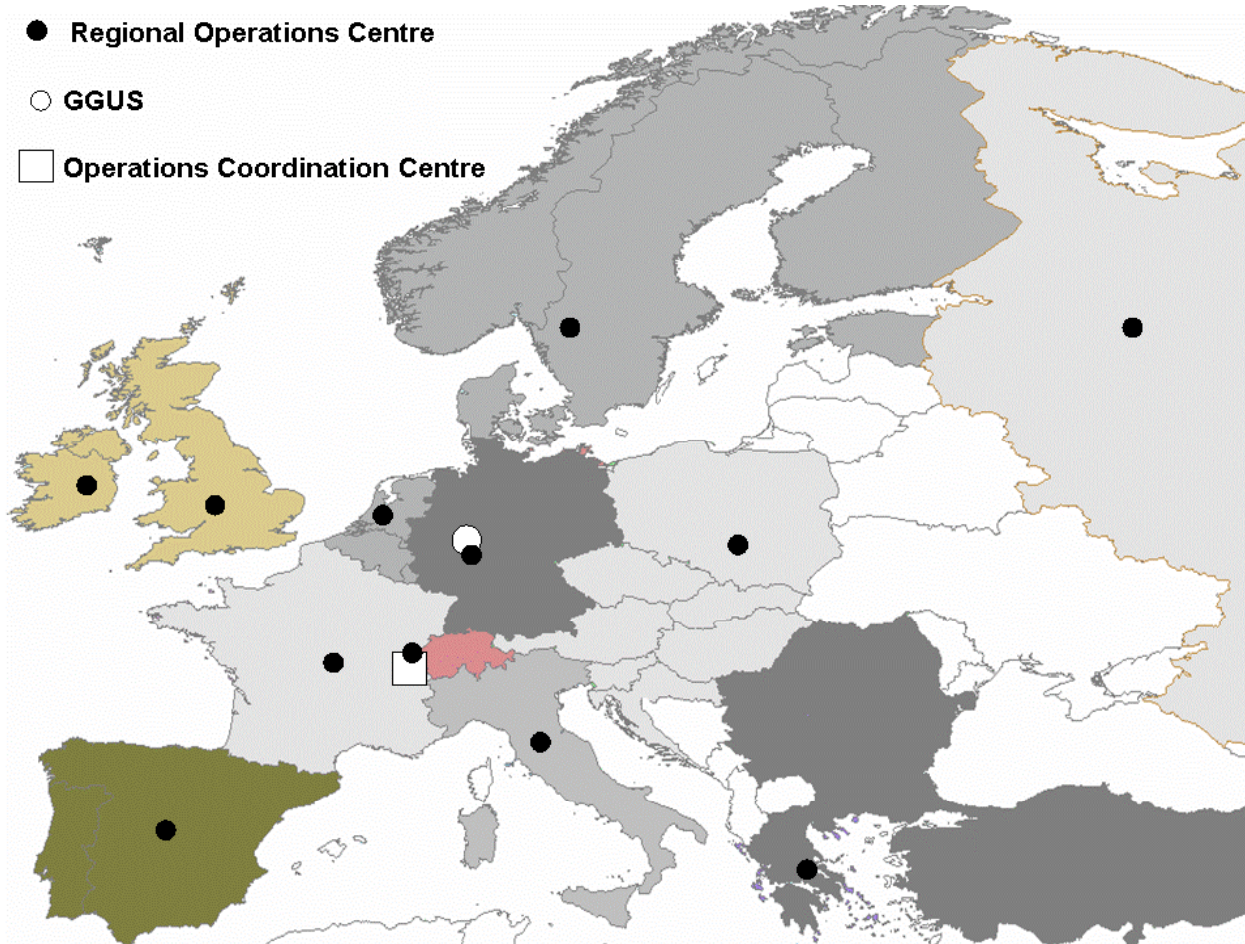
Current production middleware



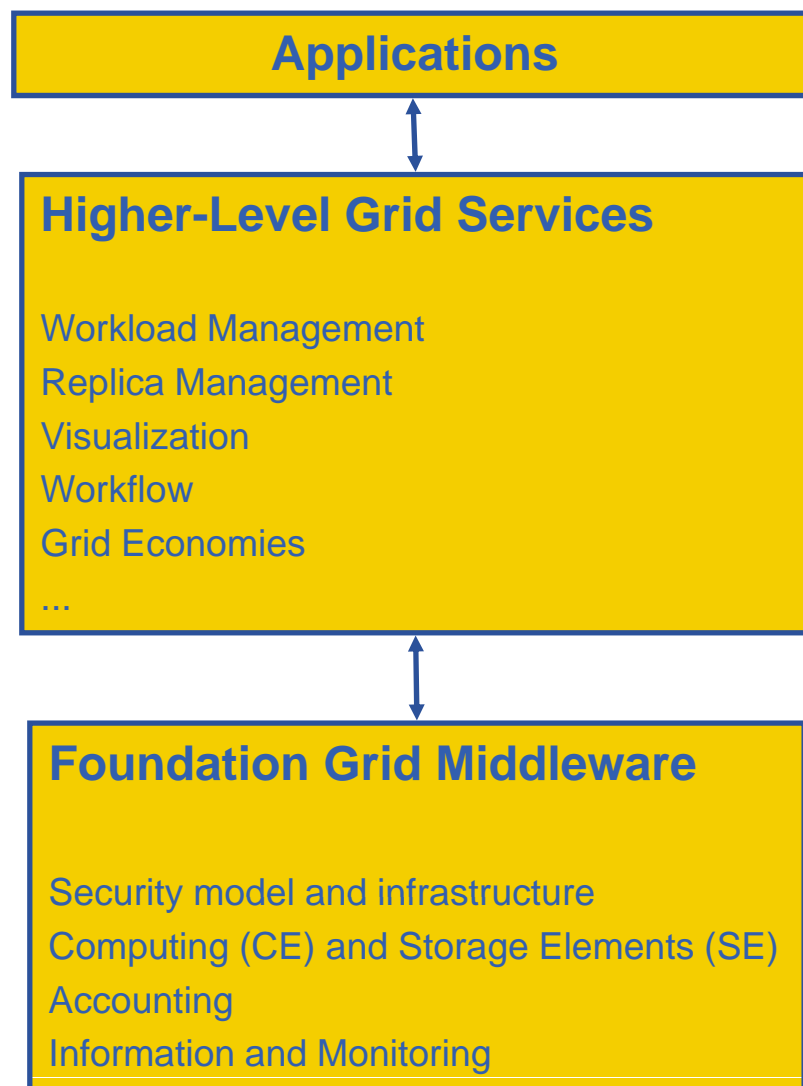
Who provides the resources?!

<u>Service</u>	<u>Provider</u>	<u>Note</u>
<u>Access service</u>	User / institute / VO	Computer with client software
<u>Resource Broker (RB)</u>	VOs - EGEE does not fund RBs	
<u>Information System:</u>	Grid operations	
<u>Computing Element (CE)</u>	VOs - EGEE does not fund CEs	Scalability requires that VOs provide resources to match average need
<u>Storage Element (SE)</u>	VOs - EGEE does not fund SEs	
"VO": virtual organisation		"Grid operations": funded effort

- Regional Operations Centre
- GGUS
- Operations Coordination Centre

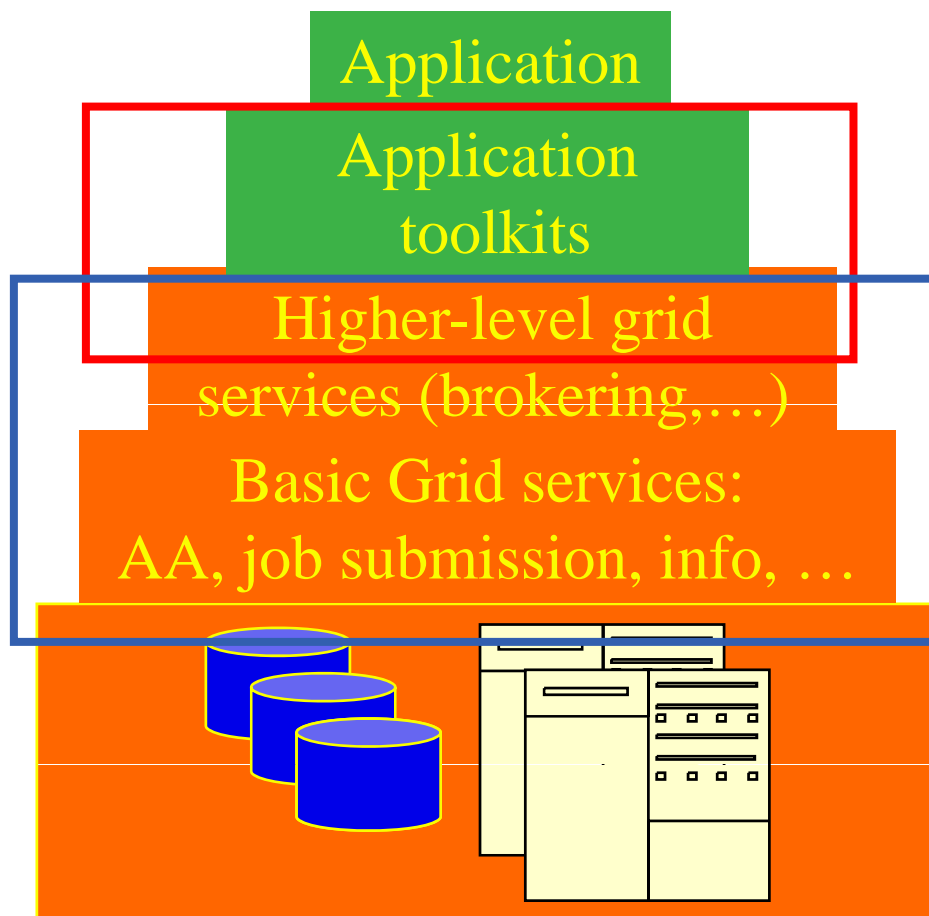


- **Operations Coordination Centre (OCC)**
 - management, oversight of all operational and support activities
- **Regional Operations Centres (ROC)**
 - providing the core of the support infrastructure, each supporting a number of resource centres within its region
 - **Grid Operator on Duty**
- **Resource centres**
 - providing resources (computing, storage, network, etc.);
- **Grid User Support (GGUS)**
 - At FZK, coordination and management of user support, single point of contact for users



- **Access for applications to:**
 - Higher-level Grid Services
 - Foundation Grid Middleware
- **Higher-Level Grid Services are supposed to help the users building their computing infrastructure but should not be mandatory**
- **Foundation Grid Middleware will be deployed on the EGEE infrastructure**
 - Must be complete and robust
 - Should allow interoperation with other major grid infrastructures
 - Should not assume the use of Higher-Level Grid Services

Overview paper <http://doc.cern.ch/archive/electronic/egee/tr/egee-tr-2006-001.pdf>



Where computer science meets the application communities!

High level toolkits and services:

- Portals – P-GRADE
- Job management – Ganga
- Alternative WMS - GridWAY
- Workflow
- Semantics, ontologies
- Registries of VO services

Production grids provide these services.

- **Many differing application domains**
 - **High-Energy Physics:** LHC, Tevatron, HERA, ...
 - **Biology:** Medical Images, Bioinformatics, Drug Discovery
 - **Earth Science:** Hydrology, Pollution, Climate, Geophysics, ...
 - **Astrophysics:** Planck, MAGIC
 - **Fusion**
 - **Computational Chemistry**
 - **Related Projects:** Finance, Digital Libraries, ...
 - **New areas:** nanotechnology, ...



- **Simulation**
 - Large number of CPU-intensive, independent jobs
 - Small input; large output
 - Run by few (expert) users
- **Bulk Processing**
 - Widely-distributed input data
 - Significant amount of input and output data
 - Meta-data services
 - More sophisticated data management
- **Responsive Apps.**
 - Small amounts of input and output data
 - Not CPU-intensive
 - Short response time (few minutes) / Rapid response: a human waiting for the result!
 - No real support in gLite → Interactive Grid – FP6 project

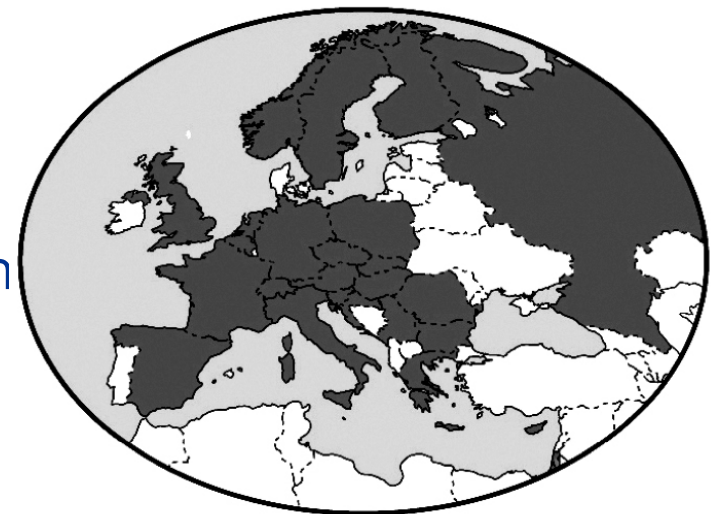
- **Workflow**
 - Complex set of algorithms for the analysis
 - Complex dependencies between individual tasks
 - No support in gLite yet
 - Higher level tools are needed (e.g. P-GRADE, Moteur)
- **Parallel Jobs**
 - Many interdependent, communicating tasks
 - Many CPUs needed simultaneously
 - Use of MPI libraries
 - Limited support in gLite: MPI configuration is not uniform
- **Legacy Applications**
 - Licenses: control access to software on the grid
 - No recompilation \Rightarrow no direct use of grid APIs!

- **Application Identification and Support (NA4)**
 - 25 countries, 40 partners, 280+ participants, 1000s of users

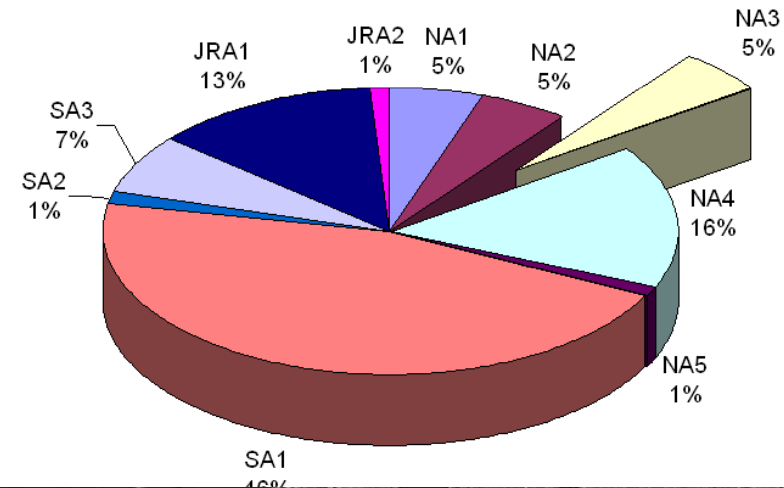
- **Support the large and diverse EGEE user community:**
 - **Promote dialog:** Users' Forums & EGEE Conferences
 - **Technical Aid:** Porting code, procedural issues
 - **Liaison:** Software and operational requirements

- **Need active participation:**
 - **Feedback:** Infrastructure, configuration, and middleware
 - **Resources:** Hardware and human

- <http://egeena4.lal.in2p3.fr>



- Expand portfolio of training materials & courses
- Train a wide variety of EGEE users (internal/external)
- Develop effective mechanisms for training end-users of the EGEE infrastructure
- Enhance e-Learning structure and provision of t-Infrastructure
- Validate cohorts of trainers & compile directory
- Collaborate in cross-activity initiatives
- <http://www.egee.nesc.ac.uk/>



- **EGEE digital library:** <http://egee.lib.ed.ac.uk/>
- **EGEE** www.eu-egee.org
- **gLite** <http://www.glite.org>

- **EGEE is running the largest multi-VO grid in the world!**
- **Creating the “grid layer” in e-Infrastructure for research, public service and industry**
- **Key concepts for EGEE**
 - Sustainability – planning for the long-term
 - Production quality
 - User support → End users: NA3 and NA4
- **EGEE’s middleware: gLite**