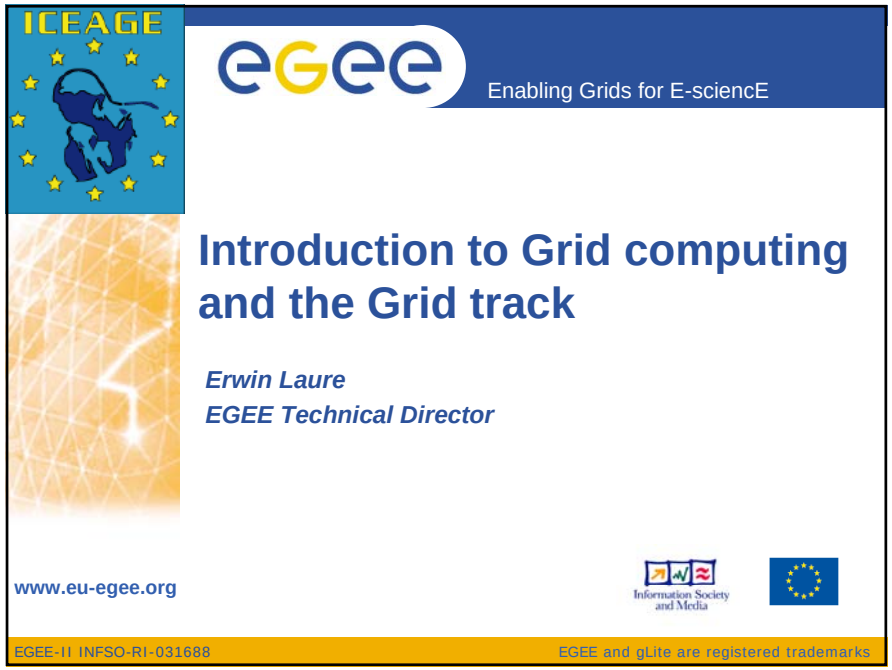
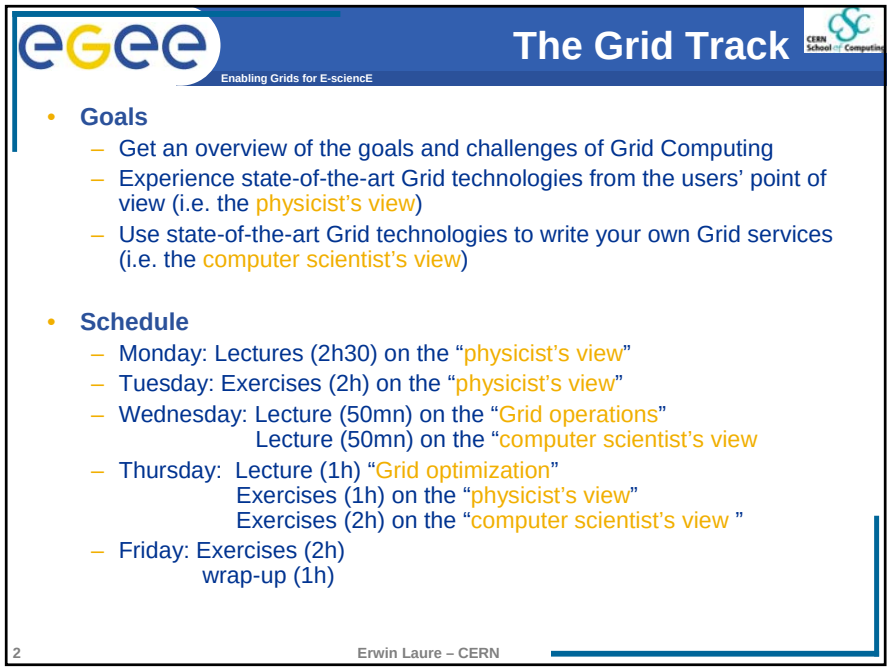

ICEAGE
Enabling Grids for E-sciencE
Introduction to Grid computing and the Grid track
Erwin Laure
EGEE Technical Director
www.eu-egee.org
Information Society and Media
EGEE-II INFSO-RI-031688
EGEE and gLite are registered trademarks


Enabling Grids for E-sciencE
The Grid Track
CERN School of Computing
Goals
Get an overview of the goals and challenges of Grid Computing
Experience state-of-the-art Grid technologies from the users' point of view (i.e. the physicist's view)
Use state-of-the-art Grid technologies to write your own Grid services (i.e. the computer scientist's view)
Schedule
Monday: Lectures (2h30) on the "physicist's view"
Tuesday: Exercises (2h) on the "physicist's view"
Wednesday: Lecture (50mn) on the "Grid operations"
Lecture (50mn) on the "computer scientist's view
Thursday: Lecture (1h) "Grid optimization"
Exercises (1h) on the "physicist's view"
Exercises (2h) on the "computer scientist's view "
Friday: Exercises (2h)
wrap-up (1h)
2
Erwin Laure – CERN

Pre-requisites

Enabling Grids for E-sciencE

- **No Grid background knowledge required**
  - PKI security infrastructure (CA's, proxies) from previous week (Alberto Pace)
- **Basics of Unix/Linux is assumed**
  - Basic knowledge of:
    - Shell scripts, perl/python, C++, Java
- **Exercises will be in teams of 2 persons**
  - At least one team member should have basic knowledge of Unix, scripting and programming languages

What is the Grid?

Enabling Grids for E-sciencE

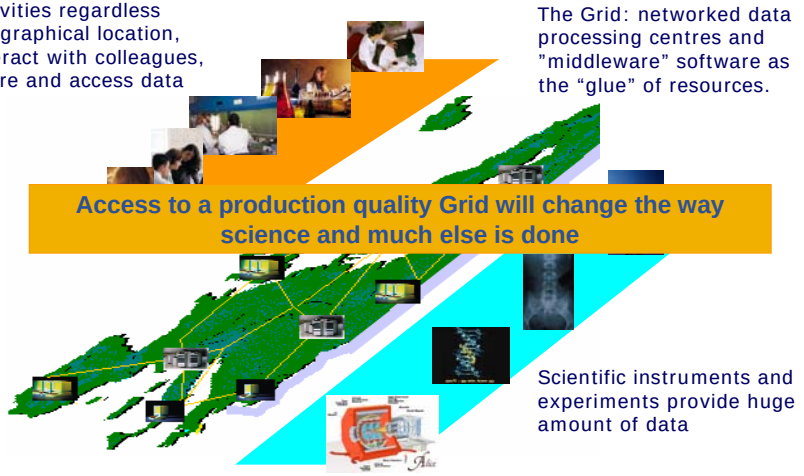
- **The World Wide Web provides seamless access to information that is stored in many millions of different geographical locations**
- **In contrast, the Grid is a new computing infrastructure which provides seamless access to computing power and data distributed over the globe**
- **The name Grid is chosen by analogy with the electric power grid: plug-in to computing power without worrying where it comes from, like a toaster**

The Grid Vision

Enabling Grids for E-sciencE

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

The Grid: networked data processing centres and "middleware" software as the "glue" of resources.



Access to a production quality Grid will change the way science and much else is done

Scientific instruments and experiments provide huge amount of data

Virtualization & Sharing

Enabling Grids for E-sciencE

- **Virtual Organizations**
  - People from different organizations but with common goals get together to solve their problems in a cooperative way – similar to a HEP experiment
- **Virtualized shared computing resources**
  - Members of VOs have access to computing resources outside their home institutions. Resource providers typically have a contract/MoU with the VO, not with the VO members
- **Virtualized shared data resources**
  - Similar to computing resources
- **Other resources may be shared and virtualized as well:**
  - Instruments, sensors, even people

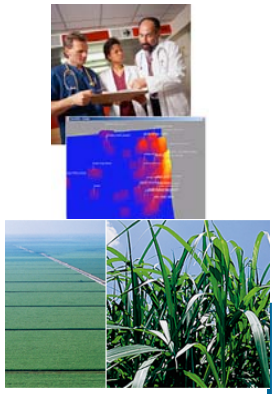
Virtualization of resources is needed to abstract from their heterogeneity

What is driving Grid development?

Enabling Grids for E-sciencE

Data and compute intensive sciences are next generation applications that have extreme needs but are likely to become mainstream in the next 5 years

- **Physics/Astronomy:** data from different kinds of research instruments
- **Medical/Healthcare:** imaging, diagnosis and treatment
- **Bioinformatics:** study of the human genome and proteome to understand genetic diseases
- **Nanotechnology:** design of new materials from the molecular scale
- **Engineering:** design optimization, simulation, failure analysis and remote Instrument access and control
- **Natural Resources and the Environment:** weather forecasting, earth observation, modeling and prediction of complex systems: river floods and earthquake simulation



How does the grid work?

Enabling Grids for E-sciencE

- The Grid relies on advanced software, called middleware, which ensures seamless communication between different computers and different parts of the world
- The Grid search engine not only finds the data the scientist needs, but also the data processing techniques and the computing power to carry them out
- It distributes the computing task to wherever in the world there is available capacity, and sends the result back to the scientist

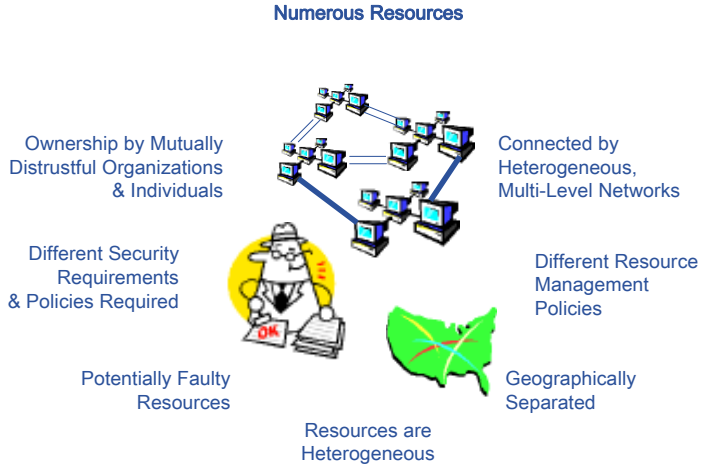
Grids vs. Distributed Computing

Enabling Grids for E-sciencE

- **Distributed applications already exist, but they tend to be *specialized systems* intended for a single purpose or user group**
- **Grids go further and take into account:**
  - Different kinds of *resources*
    - Not always the same hardware, data and applications
  - Different kinds of *interactions*
    - User groups or applications want to interact with Grids in different ways
  - *Dynamic* nature
    - Resources and users added/removed/changed frequently
- **Grids are an evolution of distributed computing techniques**

What are the characteristics of a Grid system?

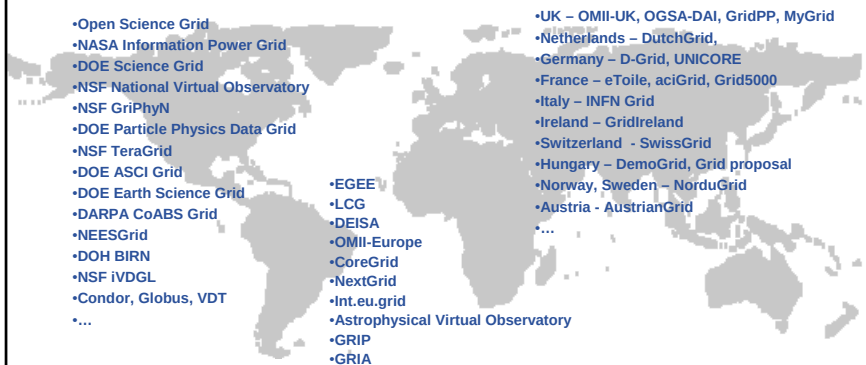
Enabling Grids for E-sciencE



Major Grid projects past and present

Enabling Grids for E-sciencE

Many Grid development efforts — all over the world



- Open Science Grid
- NASA Information Power Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF TeraGrid
- DOE ASCI Grid
- DOE Earth Science Grid
- DARPA CoABS Grid
- NEESGrid
- DOH BIRN
- NSF iVDGL
- Condor, Globus, VDT
- …

- EGEE
- LCG
- DEISA
- OMII-Europe
- CoreGrid
- NextGrid
- Int.eu.grid
- Astrophysical Virtual Observatory
- GRIP
- GRIA
- GridLab
- EGSO
- …

- UK – OMII-UK, OGSA-DAI, GridPP, MyGrid
- Netherlands – DutchGrid,
- Germany – D-Grid, UNICORE
- France – eToile, aciGrid, Grid5000
- Italy – INFN Grid
- Ireland – GridIreland
- Switzerland - SwissGrid
- Hungary – DemoGrid, Grid proposal
- Norway, Sweden – NorduGrid
- Austria - AustrianGrid
- …

Standards are the key

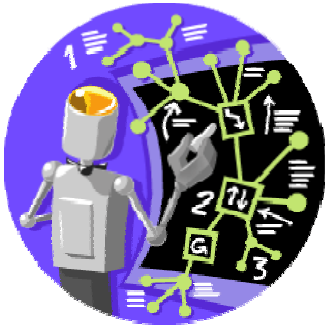
Enabling Grids for E-sciencE

- **The systems developed and deployed by the plethora of Grid projects need eventually converge**
- **Standardization efforts at multiple levels:**
- **Basic Infrastructure:**
  - OASIS (http://www.oasis-open.org/)
    Web Services, XML, WSRF
- **Higher level services**
  - OGF (http://www.ogf.org/)
    Open Grid Service Architecture (OGSA)
    over 50 Research and Working groups organized in 8 areas

What do we expect?

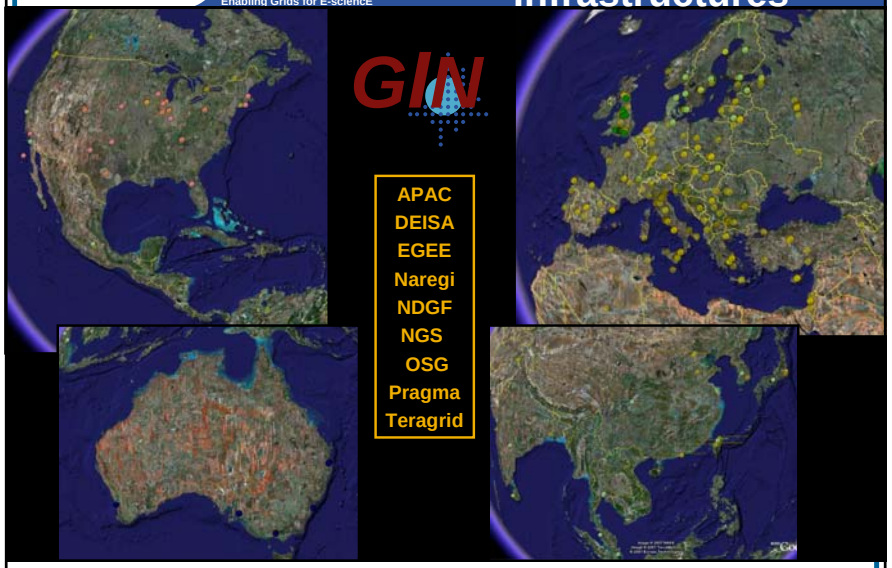
Enabling Grids for E-sciencE

- **The Grid provides:**
  - Access to a world-wide virtual computing laboratory with almost infinite resources
  - Possibility to organize distributed scientific communities in Virtual Organizations (VOs)
  - Transparent access to distributed data and easy workload management
  - Easy to use application interfaces



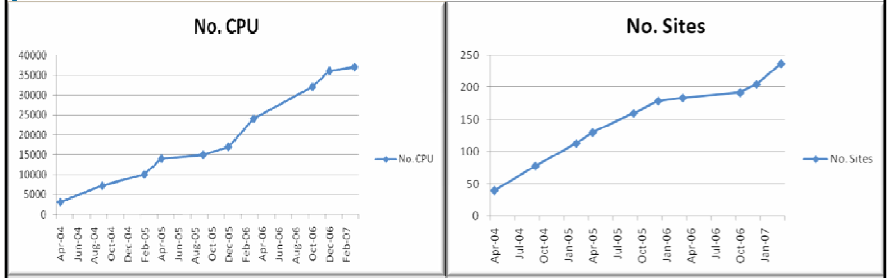
Worldwide Grid Infrastructures

Enabling Grids for E-sciencE



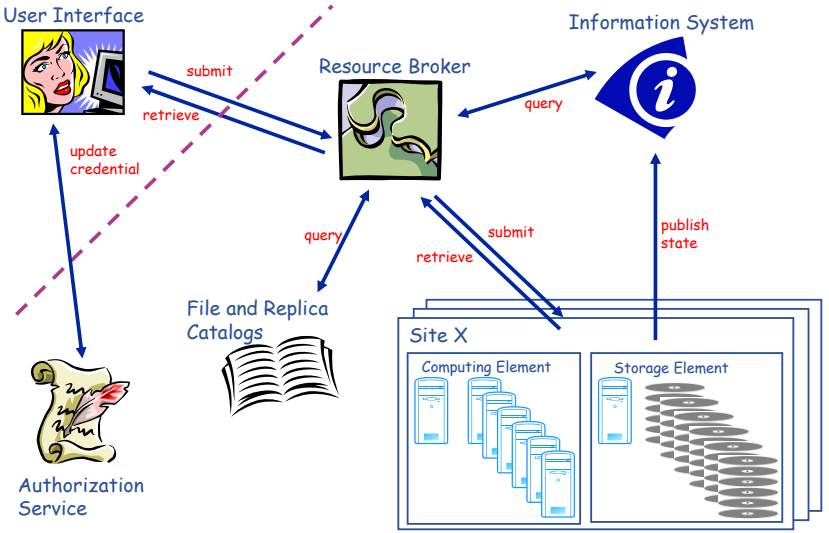
Grids are a Reality

Enabling Grids for E-sciencE



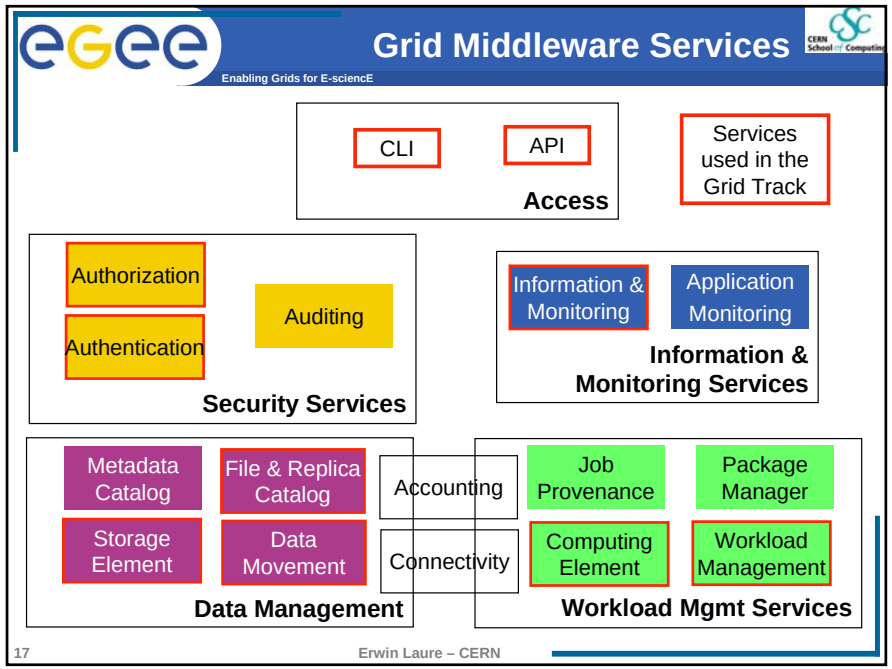
Grid Topology and Services

Enabling Grids for E-sciencE



Grid Middleware Services

Enabling Grids for E-sciencE

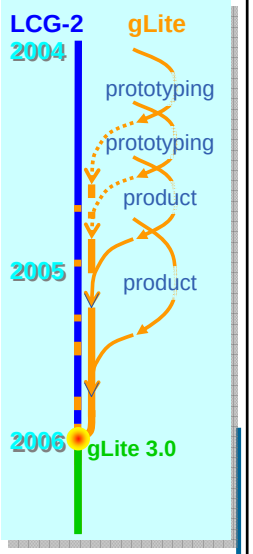


EGEE Middleware: gLite

Enabling Grids for E-sciencE

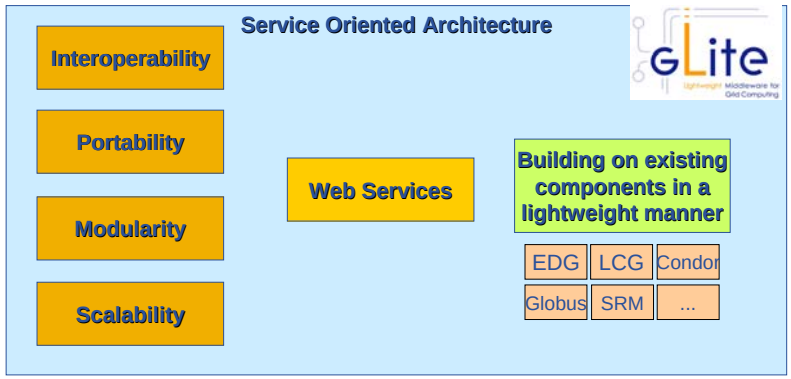
- **gLite middleware distribution re-engineered according to industry strength process**
  - Based on VDT (Condor, Globus, etc.)
- **LCG-2 distribution was baseline, new services re-engineered and successively included.**
- **Convergence with gLite 3.0**
  - Publicly released on May 4, 2006

gLite Lightweight Middleware for Grid Computing



Guiding Principles

Enabling Grids for E-sciencE



Building Grid Services

Enabling Grids for E-sciencE

- **Open Grid Service Architecture (OGSA)**
  - Extends web services to Grid services
  - Allows to build Grid services using a common infrastructure
- **First implementations appeared with Globus Toolkit version 3 (GT3) and 4 (GT4)**

This will be the focus of Thursday and Friday

Overview of Today

Enabling Grids for E-sciencE

- **The "Physicist's view" – Using the Grid**
  - Information services (20' – Heinz Stockinger)
  - Workload management (50' – Heinz Stockinger)
  - Data management (50' – Erwin Laure)
- **Exercises follow tomorrow**

More Information

Enabling Grids for E-sciencE

- **The EGEE Project**
  - http://www.eu-egee.org
- **The ICEAGE Project**
  - http://www.iceage-eu.org
- **The LCG Project**
  - http://cern.ch/lcg
- **The gLite middleware**
  - http://www.glite.org
- **The Condor Project**
  - http://www.cs.wisc.edu/condor
- **The Globus Project**
  - http://www.globus.org