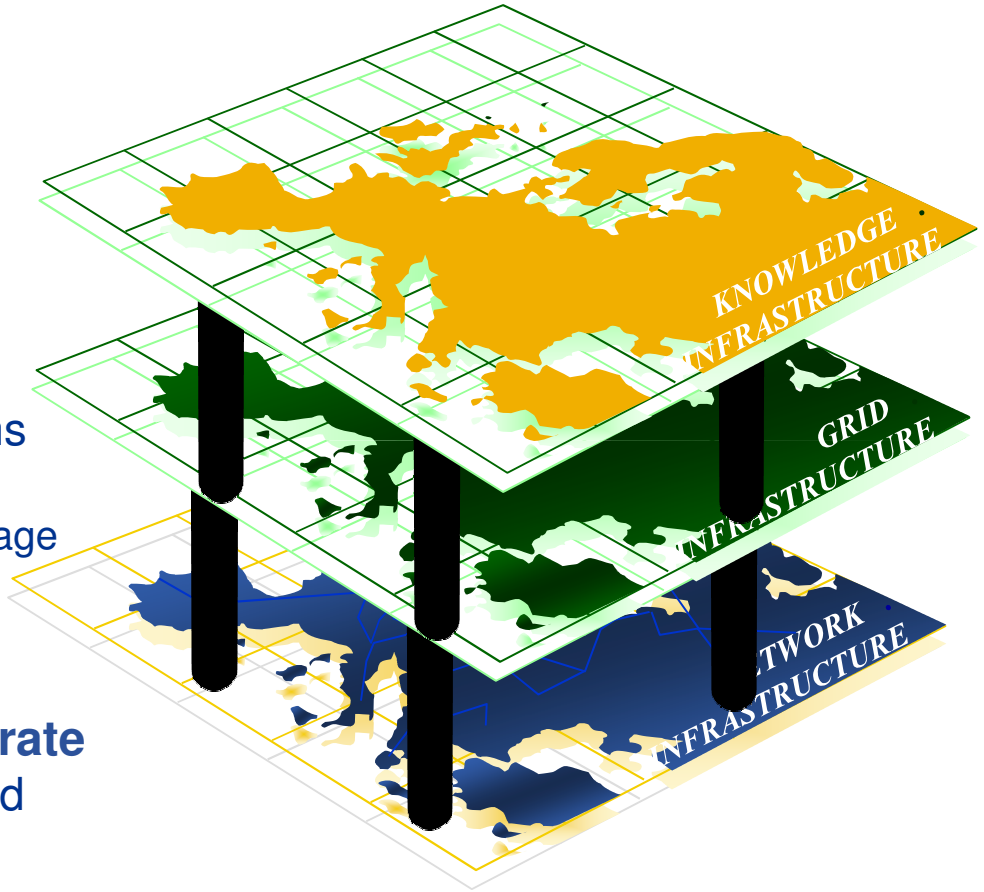
The use of grids for research infrastructures

Bob Jones
EGEE Project Director

www.eu-egee.org

How e-Infrastructures serve e-Science

- **e-Infrastructures provide easier access for**
  - Small research groups
  - Scientists from many different fields
  - Remote and still developing countries
- **To new technologies**
  - Produce and store massive amounts of data
  - Transparent access to millions of files across different administrative domains
  - Low cost access to resources
    - Mobilise large amounts of CPU & storage on short notice (PC clusters)
  - High-end facilities (supercomputers)
- **And help to find new ways to collaborate**
  - Develops applications using distributed complex workflows
  - Eases distributed collaborations
  - Provides new ways of community building
  - Gives easier access to higher education



EGEE – What do we deliver?

- **Infrastructure operation -** ***Sites distributed across many countries***
  - Large quantity of CPUs and storage
  - Continuous monitoring of grid services & automated site configuration/management
  - Support multiple Virtual Organisations from diverse research disciplines

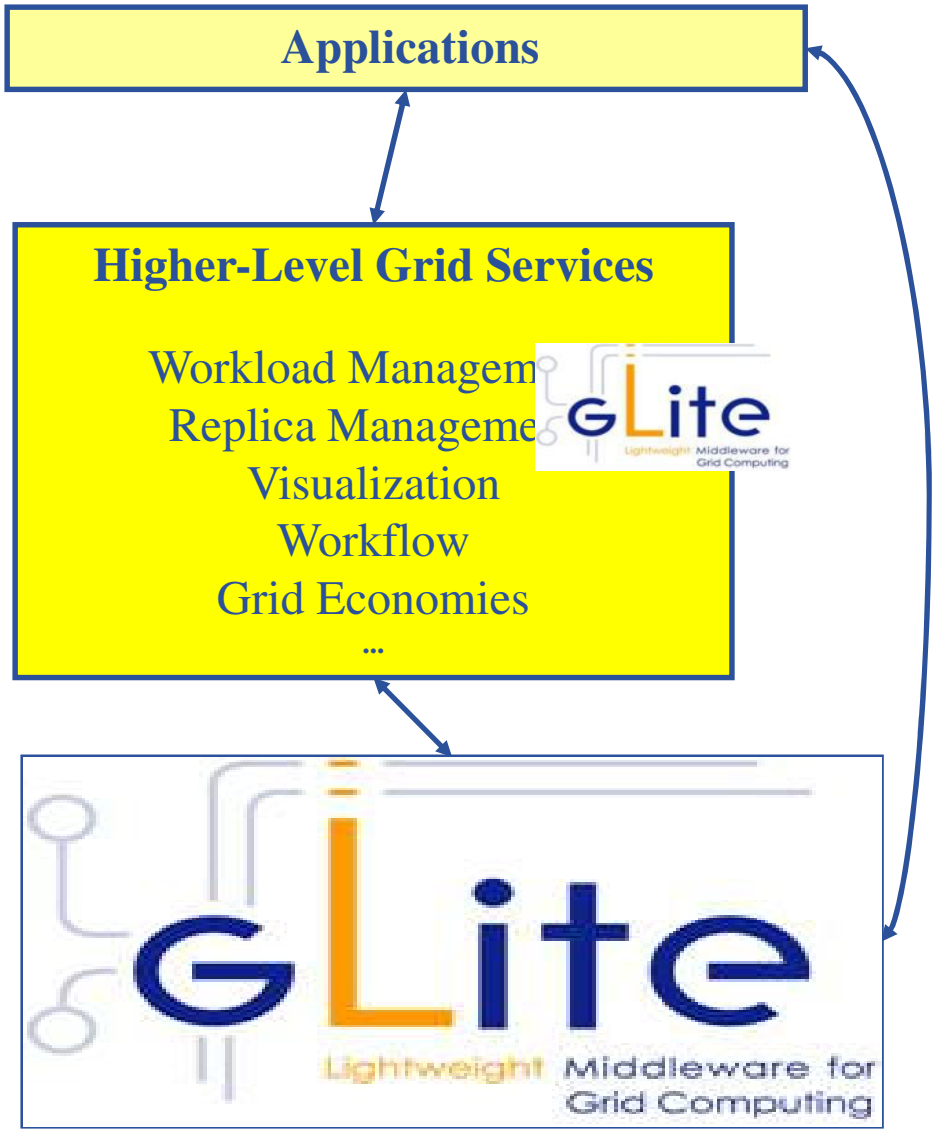
- **Middleware -** ***Production quality software distributed under business friendly open source licence***
  - Implements a service-oriented architecture that virtualises resources
  - Adheres to recommendations on web service inter-operability and evolving towards emerging standards

gLite Lightweight Middleware for Grid Computing

- **User Support -** ***Managed process from first contact through to production usage***
  - Training
  - Expertise in grid-enabling applications
  - Online helpdesk
  - Dedicated support for specific disciplines
  - Networking events (User Forum, Conferences etc.) for cross-discipline interaction

GGUS Global Grid User Support

Grid Middleware



- **Applications access both Higher-level Grid Services and Foundation Grid Middleware**
  - Application code
  - Frameworks
  - Community Portals
- **VOs complement gLite with other high-level services via the RESPECT programme**
  - Rec. External Software Pkgs. for the EGEE Community
  - Identify useful, 3rd-party software that works with gLite
  - Make users aware of that software to avoid duplicated efforts

Collaborating e-Infrastructures

Enabling Grids for E-sciencE

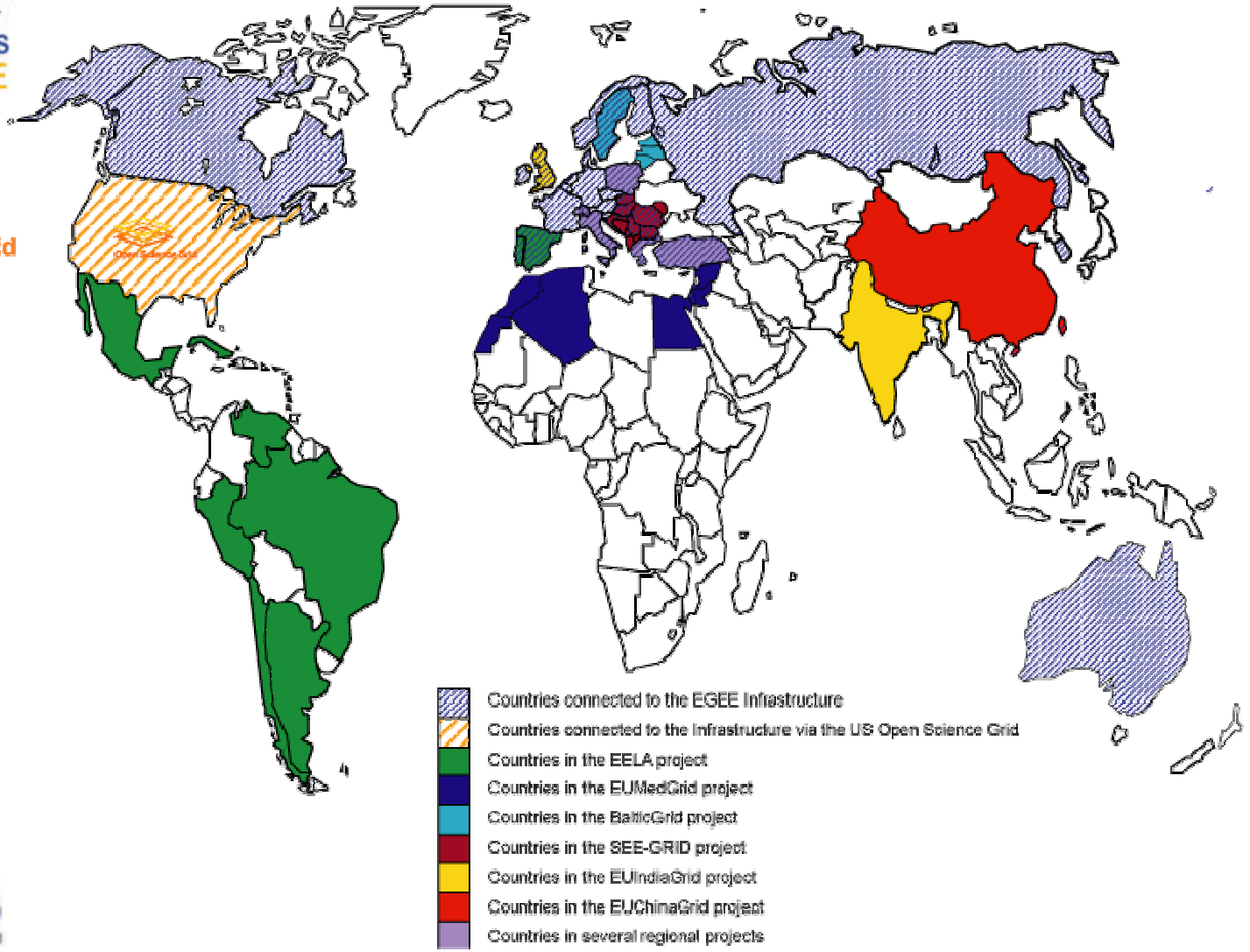
Open Science Grid

TeraGrid

Distributed European Infrastructure for Supercomputing Applications

NAREGI

GÉANT2



Lessons learnt - prototyping

- **Prototyping with selected user cases**
  - Establish a gap analysis and determine priorities
  - Investigate most effective way in which infrastructure could be extended in accordance with these priorities
  - Engage effort via collaborators for application porting, custom service development, user training etc.

LIFEWATCH

European XFEL

FAIR Facility for Antiproton and Ion Research

GSI

Lessons learnt - collaboration

- **Ensure engagement of the scientific collaboration**
  - For many scientific instruments, the computing requirements are not taken into account in the budget
  - Need to rely on contributions (personnel, resources) from collaborating partners
  - Cross-fertilisation of knowledge and tools between user communities
    - **RESPECT** collection of high-level tools and services that work on EGEE:

      http://technical.eu-egee.org/index.php?id=290

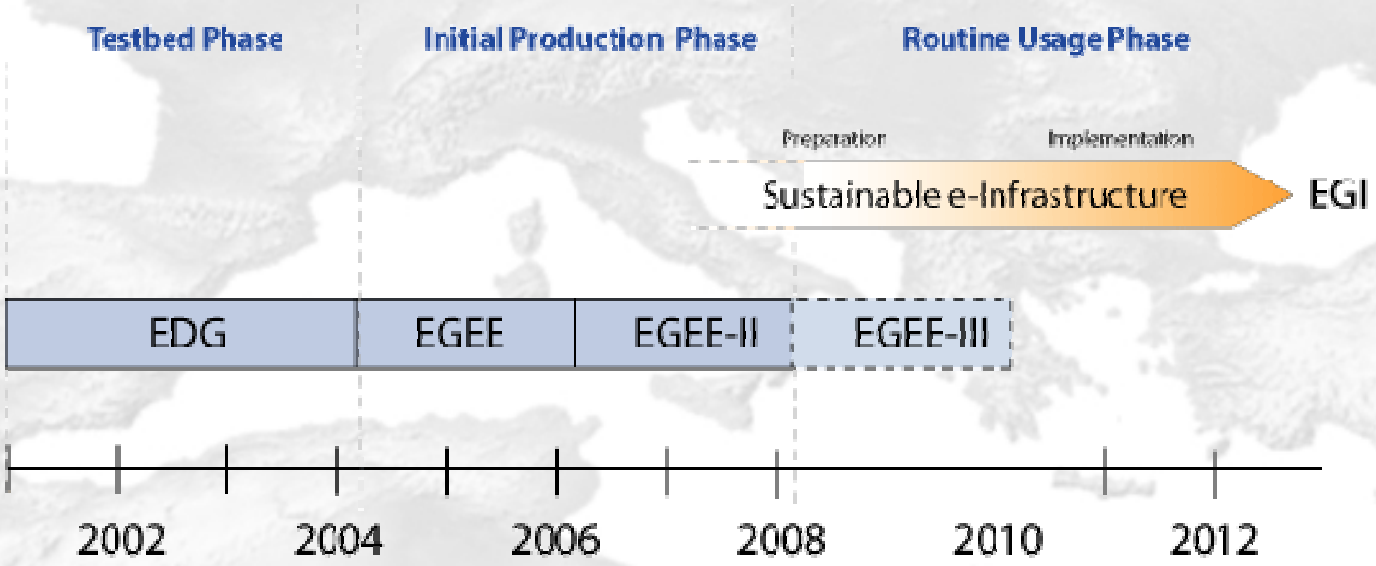
      - *Secure (encrypted) data management, metadata management, interactive applications, application lifetime management and monitoring, sensor integration*
  - Community portal to simplify access and limit training requirements
  - Need to have a long-term plan to ensure continued support
  - Get the commitments written down and monitor them!

Goal: Long-term sustainability of grid infrastructures in Europe

Approach: Establish a federated model bringing together National Grid Infrastructures (NGIs) to build the European Grid Infrastructure (EGI)

EGI Organisation: Coordination and operation of a common multi-national, multi-disciplinary Grid infrastructure

- To enable and support international Grid-based collaboration
- To provide support and added value to NGIs
- To liaise with corresponding infrastructures outside Europe



Research infrastructures and EGI

- **The creation of National Grid Infrastructures and their overall coordination can provide an ICT context for the research infrastructures**
  - An operational context for sites & centres involved in ESFRI projects
- **In the EGI context, Specialised Support Centres (SSCs) are the means of interaction with user communities**
  - The EGI SSCs are established and governed by the user communities
- **The EGI operational model and SSCs are a candidate mechanism for ESFRI preparation projects to interact with EGI**

How to be future proof

- **Consider ALL (production grids, supercomputers, commercial cloud systems, volunteer grids, network etc.) as a combined *e-Infrastructure***
  - Aim for interoperability and combine the resources into a consistent whole
  - Work closely with EGEE/EGI, DEISA/PRACE and GEANT – they are ready to help! - *they have links around the world*
- **Keep the applications agile**
  - Don't make the code so specialised that it can only use one specific installation – *things will change!*
- **Make it easy for the users**
  - Consider a community gateway/portal
    - Simplify authorisation/authentication (e.g. obtain X509 certs etc.)
    - Easy access to common codes (handle license issues)
    - Relevant tutorials & documentation

Summary

The key added value of grid infrastructures is a framework for collaboration

- Global secure access to computing resources, data, software and results
  - CPU power for computing-intensive tasks
  - Data management capabilities
    - Metadata and annotation
    - Security
    - Replication
    - High-speed data transfers
    - Facilitate creation of distributed data repositories, data mining, indexing and search
  - Software services
    - Availability of open source software
    - Integration with commercial software packages
- Scalable and dynamic architecture which can be extended with additional services as required
- All organisations can participate AND contribute

DISCUSSION

Discussion primer

These slides are purely a means to drive the discussion forward – based on the material presented yesterday

ESFRI @ EGEE09

CLARIN
Common Language Resources and Technology Infrastructure

DARIAH
Digital Research Infrastructure for the Arts and Humanities

metadata
accessing data
data harmonisation
ethics
cessda
social science data
confidentiality
data catalogue
data preservation

EPOS
EUROPEANPLATEOBSERVINGSYSTEM

EISCAT

LIFEWATCH

SKA
SQUARE KILOMETRE ARRAY

ELIXIR EUROPEAN LIFE SCIENCES INFRASTRUCTURE FOR BIOLOGICAL INFORMATION

European XFEL

Cherenkov Telescope Array

FAIR Facility for Antiproton and Ion Research
GSI

What have we seen here?

- **ESFRI projects from all sectors (with the exception of energy) represented at EGEE09**
- **Common recurring themes have been identified by the ESFRI projects**
  - Technical requirements
  - Organisational aspects
- **Everyone is aware that these are evolving plans and subject to change**

Technical requirements

- **A number of common technical requirements emerged from the ESFRI presentations:**
  - Single-sign on
  - AAA
  - Persistent, guarantee storage
  - Adherence to relevant standards
  - Support for work-flows
  - Virtual labs/organisations
  - etc.

Organisational aspects

- **A number of common organisational aspects emerged from the ESFRI presentations:**
  - Large collaboration of many institutes across many countries
  - Distributed access to data and results
  - Multiple overlapping groups requiring secure/selective access
  - No additional bureaucracy tolerated by the users
  - Science gateways as means of providing access
  - Convincing collaborations of the need for e-Infrastructures and their added value

Version: final
Date: 7th July 2009

European e-Infrastructure Forum

This document outlines the mission of the European e-Infrastructure Forum.

Mission

The European e-Infrastructure Forum is a forum for the discussion of principles and practices to create synergies for distributed Infrastructures. The goal of the European e-Infrastructure Forum is the achievement of seamless interoperation of leading e-Infrastructures serving the European Research Area. The focus of the forum is the needs of the user communities that require services which can only be achieved by collaborating Infrastructures.

The added value of this forum is that by explaining, sharing and aligning their policies the members can learn from each other and improve the services they offer to their users. As a result, the users will be able to make use of the range of resources offered by Europe's e-Infrastructures with the minimum of technical and administrative barriers. Each e-Infrastructure expects to be able to offer a broader range of facilities to users due to closer interaction of interoperable Infrastructures.

The objectives of the forum are to:

- Share information within the group on status and direction of e-Infrastructures
- Gather input and form common opinion on subjects that will impact the long-term goals, function and effectiveness of Europe's e-Infrastructures
- Present a common view on e-Infrastructures to peer groups elsewhere in the world
- Provide input on the vision and model for Europe's e-Infrastructures
- Be able to offer a clear added-value to key user communities by offering them access to Europe's e-Infrastructures in a coordinated manner and help them devise computing models that make use of the European e-Infrastructures
- Interact with emerging structures to transmit the experience gained from existing production Infrastructures and formulate common user requirements

The membership of the forum is limited to representatives of large-scale, multi-national, multi-disciplinary Infrastructures. New members of the forum will be invited to join subject to agreement by a majority of the existing members. The initial membership is drawn from the following Infrastructures:

- EGEE
- EGI
- DEISA
- PRACE
- Terena
- GEANT

Could this forum be useful to the interaction between ESFRI projects and e-Infrastructures?

Proposed next steps (1)

- **Identify clear contact points between the ESFRI projects and e-Infrastructures**
  - E-Infrastructure projects have been talking to *individuals (users or partners)*
    - *Can we make contacts more official and identify contact points in specific areas:*
      - *Security*
      - *Data management*
      - *Network*
      - *Etc.*
  - These will be useful for establishing links between different ESFRI projects, between ESFRI projects an e-Infrastructures etc.

Proposed next steps (2)

- **Use these contacts to build matrix for technical requirements & organisational aspects**

requirement	CLARIN	DARIAH/ CESSDA	EISCAT3D	EPOS	LIFEWATCH	ELIXIR	XFEL	CTA	FAIR	SKA
Single sign-on	X	X		X	X	X	X	X		
Persistent storage	X	X	X	X	X	X	X	X		X
Global		X		X				X	X	X
workflows	X				X	X		X		X
Virt Org	X				X		X	X		
stds	X	X			X					

Proposed next steps (3)

- **Once the matrix has been built it can be used to focus:**
  - Collaboration between ESFRI projects
  - Collaboration between ESFRI projects and e-Infrastructures
  - Provide input to roadmaps for e-Infrastructures of the future
  - Provide input to national funding agencies and European Commission on their future funding programmes