Digital repositories and Grids

Bob Jones
EGEE-II Project Director
CERN
Bob.Jones@cern.ch

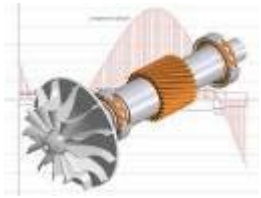
3rd July 2006
Oxford

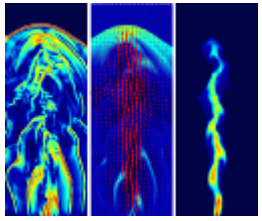
www.eu-egee.org

Information Society and Media

The project

- **EGEE**
  - 1 April 2004 – 31 March 2006
  - 71 partners in 27 countries, federated in regional Grids
- **EGEE-II**
  - 1 April 2006 – 31 March 2008
  - 91 partners in 32 countries
- **Objectives**
  - Large-scale, production-quality grid infrastructure for e-Science
  - Attracting new resources and users from industry as well as science
  - Maintain and further improve “gLite” Grid middleware





EGEE – What do we deliver?

- **Infrastructure operation**
  - Currently includes >200 sites across 39 countries
  - Continuous monitoring of grid services & automated site configuration/management
    http://gridportal.hep.ph.ic.ac.uk/rtm/launch_frame.html

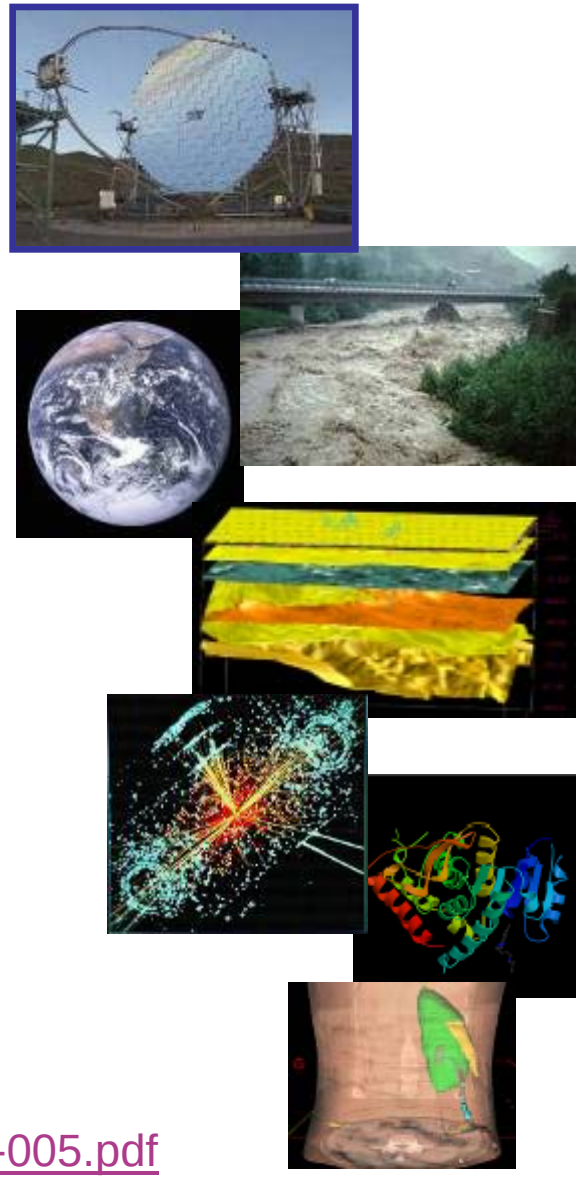
- **Middleware**
  - Production quality middleware distributed under business friendly open source licence

- **User Support** - *Managed process from first contact through to production usage*
  - Training
  - Expertise in grid-enabling applications
  - Online helpdesk
  - Networking events (User Forum, Conferences etc.)

- **Interoperability**
  - Expanding geographical reach and interoperability with related infrastructures

Applications on EGEE

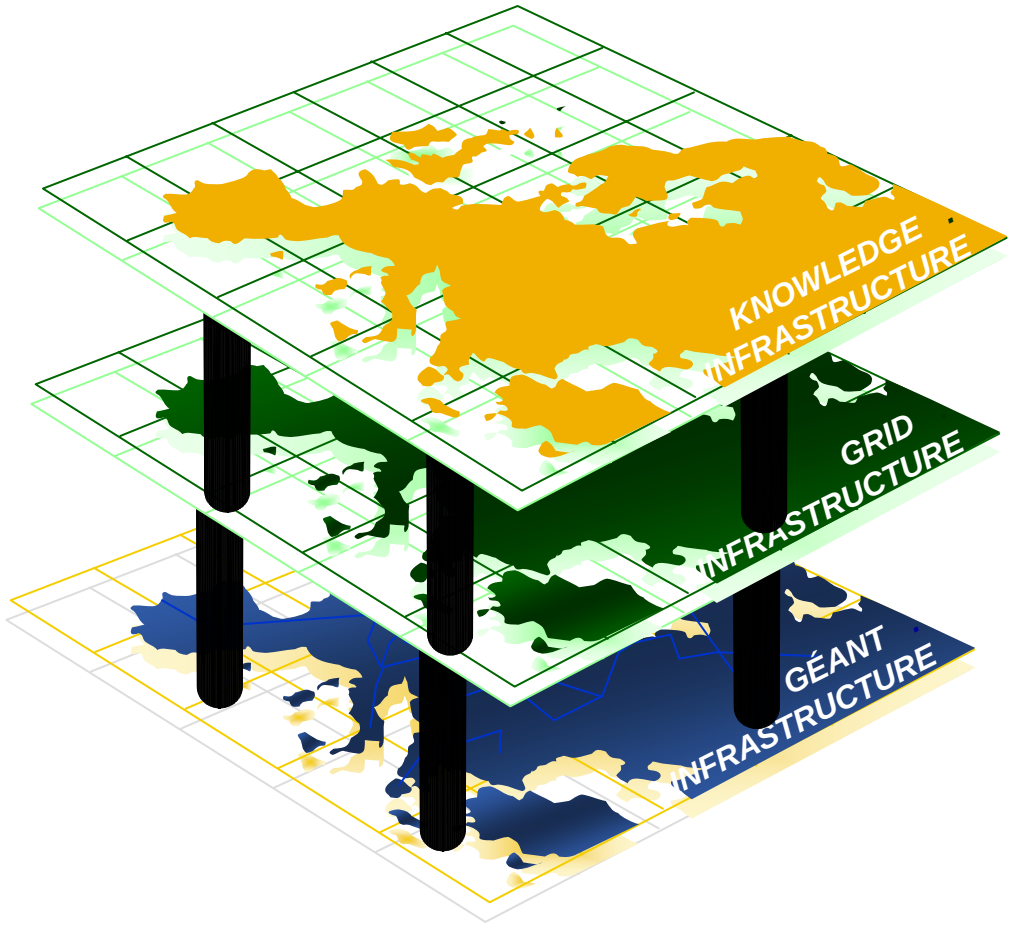
- **More than 25 applications from many domains**
  - Archaeology
  - Astrophysics
  - Computational Chemistry
  - **Digital libraries**
  - Earth Sciences
  - Financial Simulation
  - Fusion
  - Geophysics
  - High Energy Physics
  - Life Sciences
  - Multimedia
  - Material Sciences
  - …..



Book of abstracts:
http://doc.cern.ch//archive/electronic/egee/tr/egee-tr-2006-005.pdf

Support for digital repositories

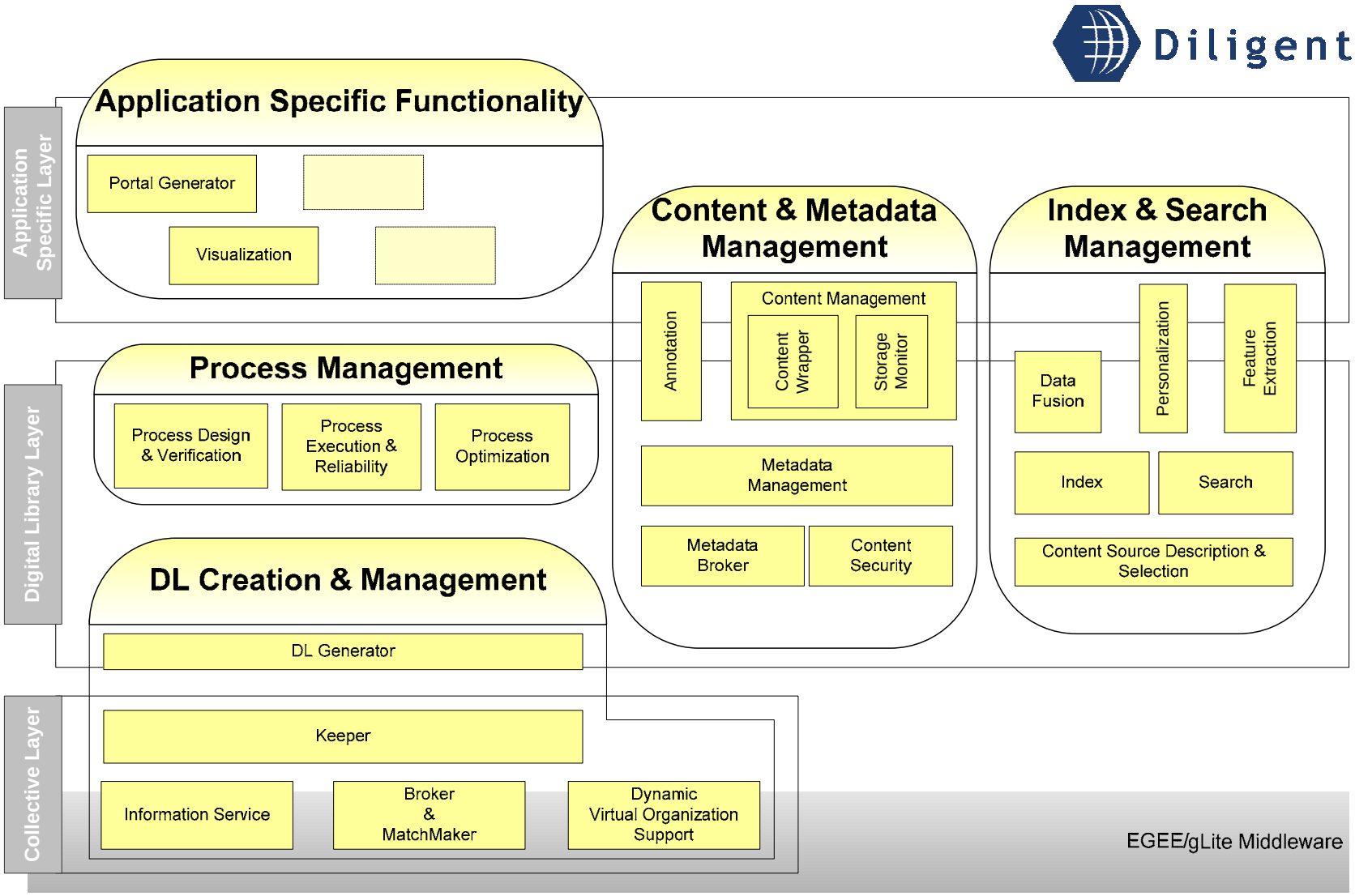
3 layered model to support access to heterogeneous information and connect resources through common shared services



Grids can offer:

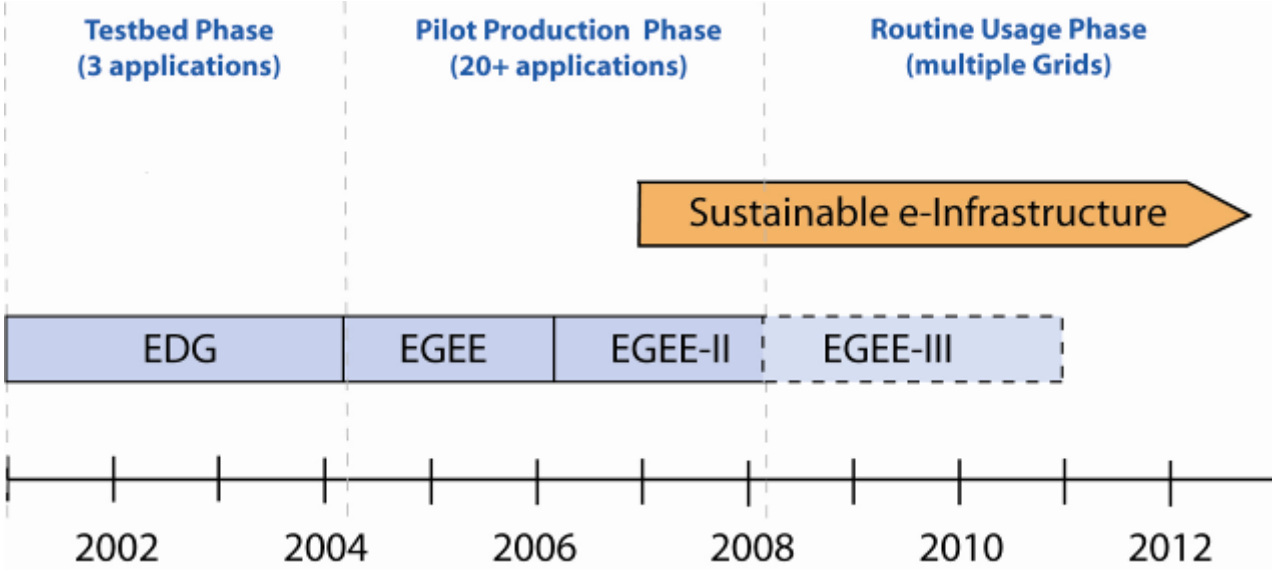
- Sharing of resources
- Secure Access Control
- Data management
- Execution of computationally demanding applications (e.g. multi-media content)

A SW architecture for digital libraries



Sustainability: Beyond EGEE-II

- **Need to prepare for permanent Grid infrastructure**
  - Maintain Europe's leading position in global science Grids
  - Ensure a reliable and adaptive support for all sciences
  - Independent of short project cycles
  - Modelled on success of GÉANT
    - Infrastructure managed in collaboration with national grid initiatives



Structure

- Federated model bringing together National Grid Initiatives (NGIs) to build a European organisation (EGI)
- EGEE-II federations would evolve into NGIs
- Each NGI is a national body
  - Recognised at the national level
  - Mobilises national funding and resources
  - Contributes and adheres to international standards and policies
  - Operates the national e-Infrastructure
  - Is application independent, open to new user communities and resource providers

European National Grid Projects

- Austria – AustrianGrid
- Belgium – BEGrid
- Bulgaria – BgGrid
- Croatia – CRO-GRID
- Cyprus – CyGrid
- Czech Republic- METACentre
- *Denmark ?*
- *Estonia – Estonian Grid*
- *Finland*
- *France – planned (ICAR)*
- Germany – D-GRID
- Greece - HellasGrid
- *Hungary*
- Ireland - Grid-Ireland
- Israel – Israel Academic Grid
- *Italy - planned*
- *Latvia – Latvian Grid*
- Lithuania - LitGrid
- Netherlands – DutchGrid
- Norway – NorGrid
- *Poland – Pioner?*
- Portugal – launched April'06
- Romania – RoGrid
- Serbia – AEGIS
- *Slovakia*
- Slovenia - SiGNET
- *Spain – planned*
- Sweden – SweGrid
- *Switzerland - SwissGrid*
- Turkey – TR-Grid
- *Ukraine - UGrid*
- United Kingdom - eScience

EGI/NGIs and Business

- **Business model to commercially exploit the research infrastructure managed by EGI/NGIs is unclear**
  - Should not use tax payers' money to compete with commercial service providers
  - Infrastructure can be used by companies to do research at a pre-competitive stage and encourage transfer of technology
- **Likely to see transfer of technology from research to industry by adoption/internalisation of EGI/NGI backed products and services (e.g. software, operations procedures/techniques)**
  - e-Science leading to e-Business
  - For multi-site corporate usage or to offer a service to a set of SMEs
- **EGI/NGIs could subcontract infrastructure support to industry and make use of commercial software as standards evolve**

Managed Resource Centres

- **Given the existence of such an e-Infrastructure, managed resources centres can be established:**
  - Create shared pool of resources (CPU, disk and data curation) independent of funding for specific user communities
  - Managed by the NGIs and coordinated by EGI
  - Joint capital funding from NGIs and EU
  - Pay-per-usage/storage business models to cover operational and depreciation costs
- **This would create a network of data-centre "hubs" that can ensure long-term preservation, service quality, redundancy and resilience**
  - Individual projects lose interest in preserving data once the project ends
  - Many end-user sites do not have the IT expertise and resources to ensure continued access to their content

Personal Lessons Learned

- **Everyone agrees a national/coordinated structure is important**
  - Provided it is to be run by them
- **Difficult to get people to run services and maintain quality**
  - Not sexy but essential and is the basis of your image
- **Must work with multiple application domains**
  - They will have different/conflicting requirements
- **Must dedicate effort to work closely with users**
  - You will underestimate the effort required
  - You will also need a salesforce
- **Access to data is paramount**
  - Requires significant network resources
  - Open Source/Access is essential
- **There will be different/rival systems**
  - Interoperability is something you owe to your users – you will have to work with these other systems
  - Developing new standards always takes longer than foreseen

EGEE'06 Conference

- EGEE'06 – Capitalising on e-infrastructures
  - Keynotes on state-of-the-art and real-world use
  - Dedicated business track exploring Grids' business potential
  - Demos and business/industry exhibition
  - Involvement of international community

- 25-29 September 2006
- Geneva, Switzerland
- http://www.eu-egee.org/egee06

Welcome to the key European Grid event of 2006!

Summary

- **Grids are all about sharing – they are an excellent means of working with groups around the world**
- **Grids can provide an infrastructure for digital repositories offering secure access, data management services and execution of computationally demanding tasks**
- **A network of managed data-centre "hubs" can offer long-term preservation, service quality, redundancy and resilience**
- **EGEE Infrastructure – world's largest multi-science production grid service**
- **Need to prepare the long-term**
  - EGEE, related EU projects, national grid initiatives and user communities are working to define a model for a sustainable grid infrastructure

www.eu-egee.org