Enabling Grids for E-sciencE

- EGEE

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

www.eu-egee.org
EGEE-II started 1 April 2006
- 2 years duration
- EGEE final review 23-24 May @ CERN

EGEE-II – a natural continuation of EGEE
- Emphasis on providing an infrastructure
  † increased support for applications
  † interoperate with other infrastructures
  † more involvement from Industry
- Expanded consortium
  † 91 partners
  † 11 Joint Research Units
- EGEE/EGEE-II transition meeting
  † CERN 12-13 April
EGEE – What do we deliver?

- **Infrastructure operation**
  - Currently include >200 sites across 39 countries
  - Continuous monitoring of grid services in a distributed global infrastructure
  - Automated site configuration/management

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

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

- **Future**
  - Expand on interoperability with related infrastructures
Mission

• **Infrastructure**
  – Manage and operate production e-Infrastructure open to all user communities and service providers
  – Interoperate with e-Infrastructure projects around the globe
  – Contribute to Grid standardisation and policy efforts

• **Support applications from diverse communities**
  – Astrophysics
  – Computational Chemistry
  – Earth Sciences
  – Finance
  – Fusion
  – Geophysics
  – High Energy Physics
  – Life Sciences
  – Multimedia
  – …

• **Business**
  – Forge links with the full spectrum of interested business partners to aid industrial take-up of grids

• **Disseminate knowledge about the Grid through training**
A global, federated e-Infrastructure

EGEE Infrastructure:
- 200 sites in 39 countries
- ~20,000 CPUs
- > 5 PB storage
- > 10,000 concurrent jobs per day
- > 60 Virtual Organisations
Expertise & Resources

- 32 countries
- 12 federations
- Major and national Grid projects in Europe, USA, Asia

+ 27 countries through related projects:
  - BalticGrid
  - SEE-GRID
  - EUMedGrid
  - EUCChinaGrid
  - EELA
Federated model bringing together National Grid Initiatives (NGIs) to build a European organisation

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
- Application independent, open to new user communities and resource providers
Industry and EGEE-II

- **Industry Task Force**
  - Group of industry partners in the project
  - Links related industry projects (NESSI, BEinGRID, …)
  - Works with EGEE’s Technical Coordination Group (TCG) to place industry requirements on equal footing

- **Collaboration with CERN openlab project**
  - IT industry partnerships for hardware and software development

- **EGEE Business Associates (EBA)**
  - Companies sponsoring work on joint-interest subjects
    - Technical developments
    - Market Surveys
    - Business modelling
    - Exploitation strategies
    - Transfer of know-how and services to industry

- **Industry Forum (representatives in most European countries)**
  - Led by Industry to improve Grid take-up in Industry
  - Organises industry events and disseminates grid information
    - Next event: Paris 27th April: [http://www.eu-egee.org/events/industryday/first-egee-industry-day/](http://www.eu-egee.org/events/industryday/first-egee-industry-day/)
Middleware in EGEE-II

Applications

Higher-Level Grid Services
- Workload Management
- Replica Management
- Visualization
- Workflows
- Grid economies
- etc.

Foundation Grid Middleware
- Security model and infrastructure
- Computing (CE) & Storage Elements (SE)
- Accounting
- Information providers and monitoring

- Provide specific solutions for supported applications
- Host services from other projects
- More rapid changes than Foundation Grid Middleware
- Deployed as application software using procedure provided by grid operations

- Application independent
- Evaluate/adhere to new stds
- Emphasis on robustness/stability over new functionality
- Deployed as a software distribution by grid operations
Software deployment paths

- Foundation grid middleware and selected high-level services are supported directly within EGEE
- Application and high-level services from other SW providers will be incorporated
Applications on EGEE

- More than 20 applications from 7 domains
  - High Energy Physics
    - 4 LHC experiments (ALICE, ATLAS, CMS, LHCb)
    - BaBar, CDF, DØ, ZEUS
  - Biomedicine
    - Bioinformatics (Drug Discovery, GPS@, Xmipp_Medline, etc.)
    - Medical imaging (GATE, CDSS, gPTM3D, SiMRI 3D, etc.)
  - Earth Sciences
    - Earth Observation, Solid Earth Physics, Hydrology, Climate
  - Computational Chemistry
  - Astronomy
    - MAGIC
    - Planck
  - Geo-Physics
    - EGEODE
  - Financial Simulation
    - E-GRID

Another 8 applications from 4 domains are in evaluation stage
Example: Data challenge on bird flu: biological goals

- The bird flu virus is named H5N1. H5 and N1 correspond to the name of proteins (Hemagglutinins and Neuraminidases) on the virus surface.
- Neuraminidases play a major role in the virus multiplication.
- Present drugs such as Tamiflu inhibit the action of neuraminidases and stop the virus proliferation.
- The virus keeps mutating and drug-resistant N1 variants can appear.
- The goal of the data challenge is to study in silico the impact of selected point mutations on the efficiency of existing drugs and to find new potential drugs.

Credit: Y-T Wu
A collaboration of 5 grid projects: Auvergrid, BioinfoGrid, EGEE-II, Embrace, TWGrid

Data challenge parameters:
- One docking software: autodock
- 8 conformations of the target (N1)
- 300,000 selected compounds
- 100 year CPU to dock all configurations on all compounds

Timescale:
- First contacts: March 1st 2006
- kick-off: April 1st 2006
- Targeted duration: 4 weeks
• 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 project funding cycles
  – Modelled on success of GÉANT
    ☑ Infrastructure managed in collaboration
      with national grid initiatives
Based on experience gathered during EGEE, the following key services have been found necessary for a central organisation in coordination with the NGIs:

- The information on the following slides are taken from existing EGEE structures and procedures as a means of explaining the concepts.

**Operation of Infrastructure**

- Runs Operational Coordination Centre linking a Regional Operations Centre (Point of Presence) in each NGI
- Coordinates the grid security and resource accounting
- Negotiates resources for user communities
- General management of the User Support process
- Interaction point with GEANT for grid issues
- Provides documentation for end users and resource providers
EGI Key Services (cont)

- **Middleware testing and certification**
  - Integrates middleware from other sources to produce distributions
  - Provides first-level middleware support team
  - Operates beta-test services for upcoming distributions

- **Application support**
  - No direct support but rather coordination of NGI support groups and management of overall application lifecycle (virtuous cycle)

- **Dissemination and outreach**
  - Branding, media relations, production of promotional material, websites etc
  - Event organisation
  - Public outreach & surveys - Ensuring higher level of media coverage as technology matures and becomes available to more end users
  - Representation of NGIs in international bodies and standards groups at a policy level

- **Training**
  - Aid in the formation of NGIs and their management/operations
  - Training material repository
1. Sustainable e-Infrastructure
   - Long-term cycle (up to 7 years)
   - Production/operations focus
   - Application independent
   - Open and inclusive

2. On-going developments
   - Further developments of middleware, applications, security etc.
   - Projects separate from yet coordinated with the infrastructure
   - limited duration (2 to 3 years)

Complementary streams where successful developments (2) should become part of the sustainable infrastructure (1)

Increase (2) $\rightarrow$ (1) by making the infrastructures stakeholders in R&D projects
• EGEE’06 – Capitalising on e-infrastructures
  – Demos
  – Related Projects
  – Industry
  – International community (UN organisations in Geneva etc.)

• 25-29 September 2006
• Geneva, Switzerland

• http://www.cern.ch/egee-intranet/conferences/EGEE06

Welcome to the key European Grid event of 2006!
Summary

- **EGEE Infrastructure** – world’s largest multi-science production grid service

- **EGEE-II** is the opportunity to expand on this existing base both in terms of scale and usage

- **EGEE-II** is ready and willing to strengthen links with F2 projects and suggests several concrete actions
  - A policy for relating infrastructure and R&D needs to be put in place

- **Need to prepare the long-term**
  - EGEE, related projects, national grid initiatives and user communities are working to define a model for a sustainable grid infrastructure that is independent of project cycles