eGee

Enabling Grids for E-sciencE

Introduction to Grids and the EGEE project

Bob Jones

EGEE-II Project Director

CERN

Bob.Jones@cern.ch

EGEE Industry Day
27 April 2006
Paris

www.eu-egee.org

Information Society
and Media

EGEE and gLite are registered trademarks

- **Science is becoming increasingly digital and needs to deal with increasing amounts of data**

- **Simulations get ever more detailed**
  - Nanotechnology – design of new materials from the molecular scale
  - Modelling and predicting complex systems (weather forecasting, river floods, earthquake)
  - Decoding the human genome

- **Experimental Science uses ever more sophisticated sensors to make precise measurements**
  - →Need high statistics
  - →Huge amounts of data
  - →Serves user communities around the world

Enabling Grids for E-sciencE
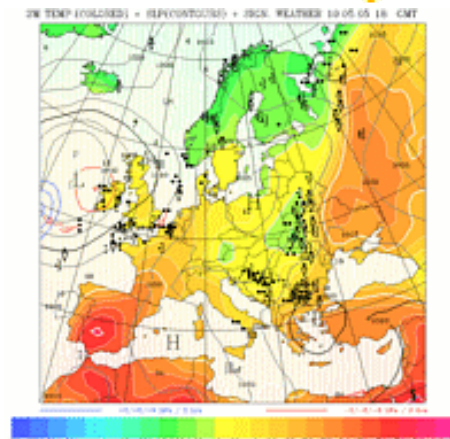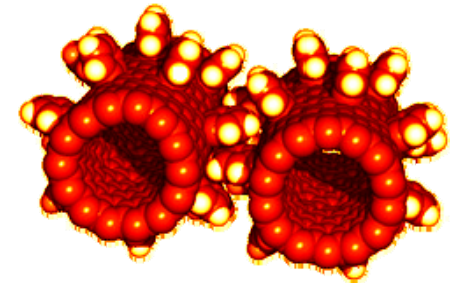
- **Integrating computing power and data storage capacities at major computer centres**

- **Providing seamless access to computing resources distributed around the globe**

→ More effective and seamless collaboration of dispersed communities, both scientific and commercial

→ Ability to run large-scale applications comprising thousands of computers, for wide range of applications
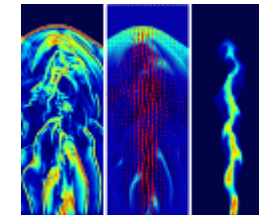
egee
Enabling Grids for E-sciencE

- **Virtual Organizations (VOs)**
  - People from different organizations but with common goals get together to solve their problems in a cooperative way – similar to a physics experiment

- **Virtualized shared computing resources**
  - Members of VOs have access to computing resources outside their home organisation. Resource providers typically have a contract/agreement with the VO, not with the VO members

- **Other resources may be shared and virtualized as well:**
  - Data, instruments, sensors

**Virtualization of resources is needed to overcome their heterogeneity**

The EGEE 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**
  - **Expanded consortium**
    - **91 partners**
    - **11 Joint Research Units**

  - **Natural continuation of EGEE**
  - **Emphasis on providing production-level infrastructure**
    - → **increased support for applications**
    - → **interoperation with other Grid infrastructures**
    - → **more involvement from Industry**

EGEE: SA 48%, NA 28%, JRA 24%

EGEE-II: SA 54%, NA 32%, JRA 14%
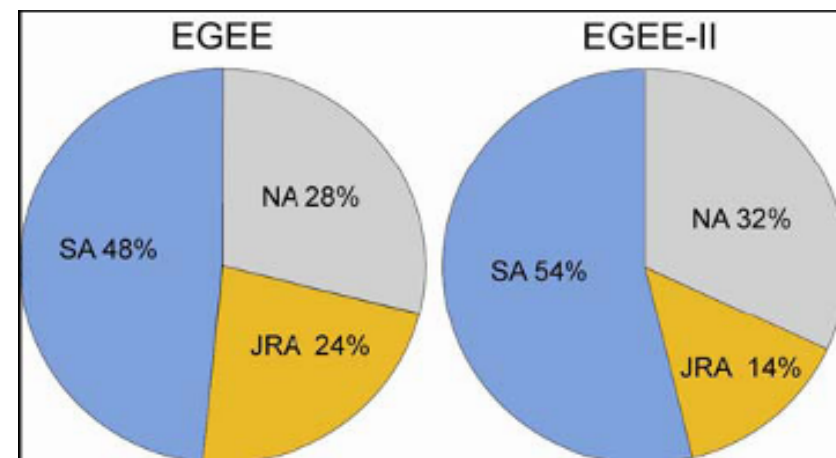
- **Infrastructure operation**
  - **Currently includes >200 sites across 39 countries**
  - **Continuous monitoring of grid services in a distributed global infrastructure**
  - **Automated site configuration/management**

- **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.)**

- **Inter-operability**
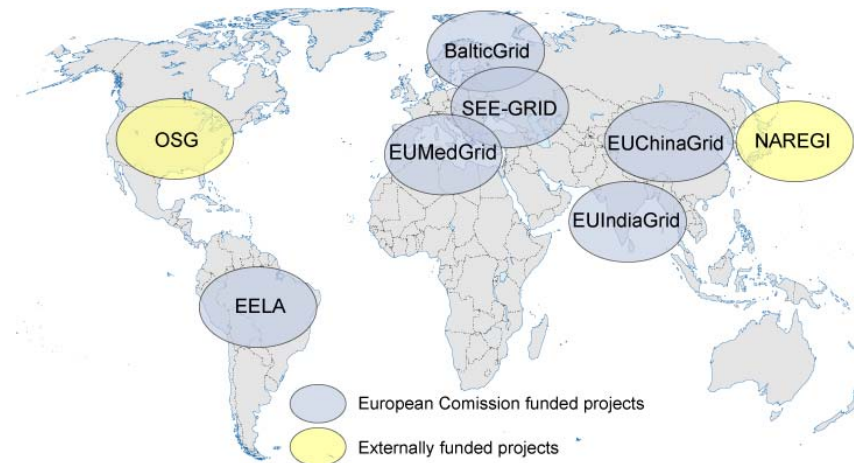 - **Expanding interoperability with related infrastructures**

Enabling Grids for E-sciencE

- **32 countries**
- **13 federations**
- → **Major and national Grid projects in Europe, USA, Asia**

+ 27 countries through related projects:

- – **BalticGrid**
- – **EELA**
- – **EUChinaGrid**
- – **EUIndiaGrid**
- – **EUMedGrid**
- – **SEE-GRID**

eGee
Enabling Grids for E-sciencE

- **More than 20 applications from many domains**
  - Astrophysics
    - MAGIC, Planck
  - Computational Chemistry
  - Earth Sciences
    - Earth Observation, Solid Earth Physics, Hyd...
  - Financial Simulation
    - E-GRID
  - Fusion
  - Geophysics
    - EGEODE
  - High Energy Physics
    - 4 LHC experiments (... AS, CMS, LHCb)
    - BaBar, CDF, DØ
  - Life Sciences
    - Bioinforma... ...covery, GPS@, Xmipp_MLrefine, etc.)
    - Medical ... ATE, CDSS, gPTM3D, SiMRI 3D, etc.)
  - Multimedia
  - …

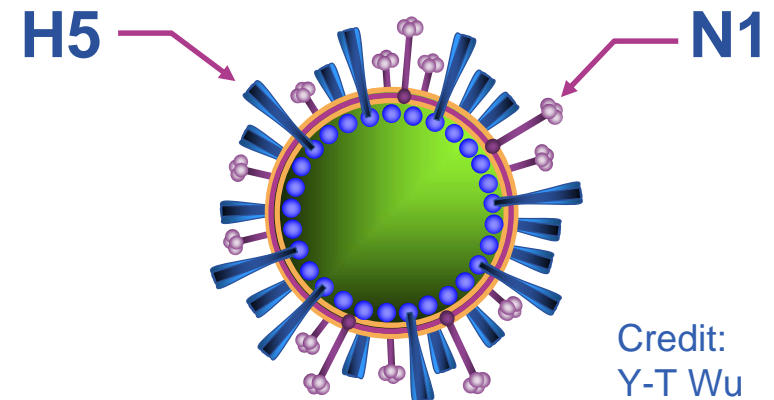**Encourage inter-disciplinary research and increase data inter-operability**
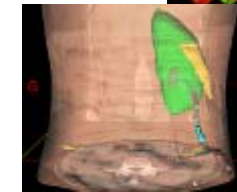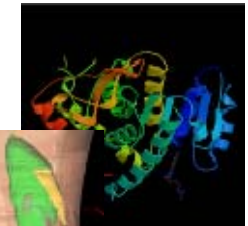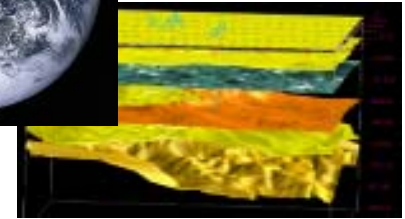
H5   N1

Credit:
Y-T Wu

- **Avian Flu H5N1**
  - H5 and N1 = proteins on virus surface (Hemagglutinins and Neuraminidases)
  - Neuraminidases play major role in virus multiplication

- **Present drugs (e.g. Tamiflu)**
  - inhibit action of neuraminidases and stop virus proliferation

- **Problems**
  - Virus keeps mutating
  - Drug-resistant N1 variants can appear

- **Biological goal of data challenge**
  - Study *in silico* the impact of selected point mutations on the efficiency of existing drugs
  - Find new potential drugs

Data challenge on avian flu (II)

- **A collaboration of 5 grid projects**
  - **Auvergrid, BioinfoGrid, EGEE, Embrace, TWGrid**

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

Credit:
Y-T Wu

- **Timescale:**
  - **First contacts established: 1 March 2006**
  - **Data Challenge kick-off: 1 April 2006**
  - **Targeted duration: 4 weeks**
  - **Status: > 75% of work under-taken with ~80% efficiency**
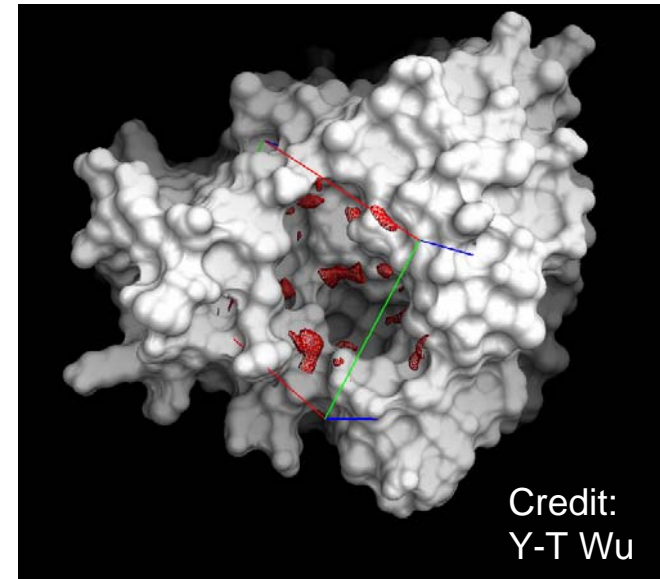
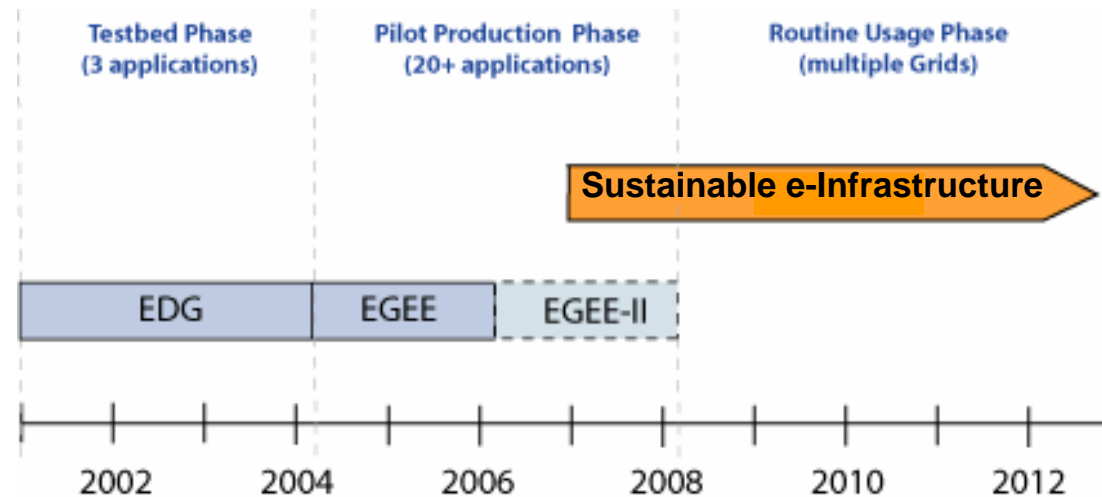**Enabling Grids for E-sciencE**

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

→ **Disseminate knowledge about the Grid trough training**

- **Trust -** Not use to sharing resources

- **Security -** Sensitive data with sensitive applications

- **Business models –** what can be charged for as a service

- **Guaranteed QoS –** Service Level Agreements

- **Accounting -** tracking resources usage in multi-admin context

- **Standards –** to encourage long-term investment

- **Applications –** need to support legacy applications

- **Portability –** across multiple platforms and implementations

- **Open source support –** robust reference implementation

- **Software license management –** how to generate revenue in a grid context

EGEE-II provides an excellent framework for collaborating with business on these subjects

- **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 funding cycles
  - Modelled on success of GÉANT
    - Infrastructure managed in collaboration with national grid initiatives

Testbed Phase (3 applications) | Pilot Production Phase (20+ applications) | Routine Usage Phase (multiple Grids)

Sustainable e-Infrastructure

EDG | EGEE | EGEE-II

2002   2004   2006   2008   2010   2012

- **Federated model**
  - National Grid Initiatives (NGIs)
  - European Organisation

- **EGEE-II Federations would evolve into NGIs**

Asia
Central Europe
CERN
France
Germany/Switzerland
Italy
Networking

Northern Europe
Russia
South-Eastern Europe
South-Western Europe
UK/Ireland
USA

- **Key Services**
  - Operation of infrastructure
  - Middleware testing and certification
  - Application support
  - Dissemination and outreach
  - Training

Enabling Grids for E-sciencE

- **EGEE Infrastructure – world's largest multi-discipline production grid service**
  - EGEE-II is the opportunity to expand on this existing base both in terms of scale and usage

- **The work of EGEE and related grid projects are critical in establishing multi-disciplinary production grids with a global scope**

- **If we succeed then the potential return to international scientific communities will be enormous and open the way for commercial and industrial applications**
  - Closer industrial involvement is actively sought

- **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

Enabling Grids for E-sciencE

- **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!