European Grid Infrastructure: Enabling the Global Research Community

Steven Newhouse
Director EGI.eu
Project Director, EGI-InSPIRE
Technical Director, EGEE-III

www.eu-egee.org

e-infrastructure

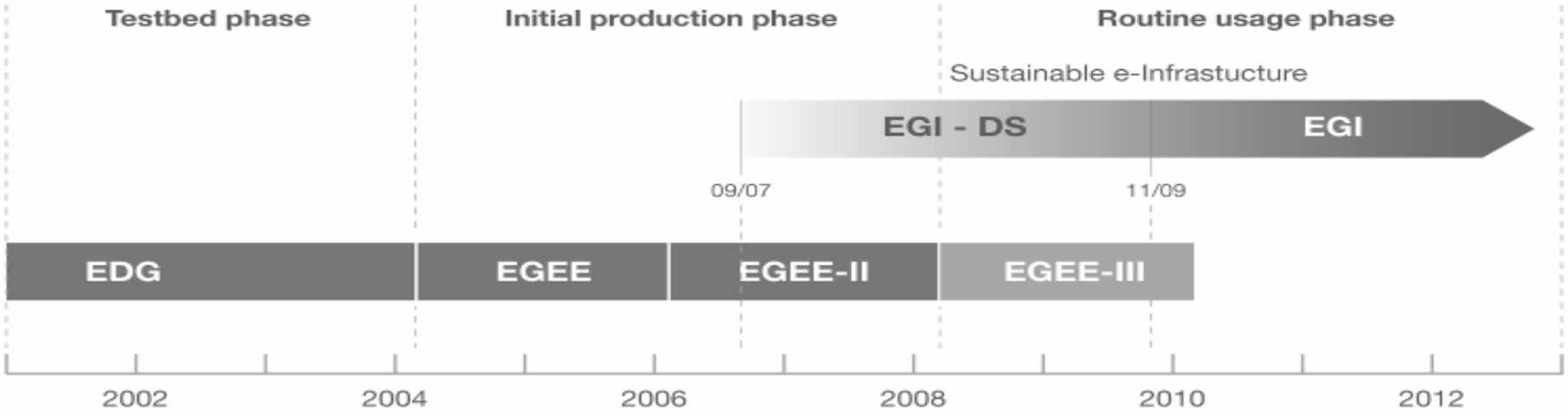
Infrastructure (Wikipedia)

- **Infrastructure** is the *basic physical and organisational structures needed for the operation of a society or enterprise*, or the services and facilities necessary for an economy to function.
- Technical structures for us are:
  - Hardware: Compute, Storage, Instruments, Sensors, …
  - Software: Authentication, Authorisation, Accounting, …

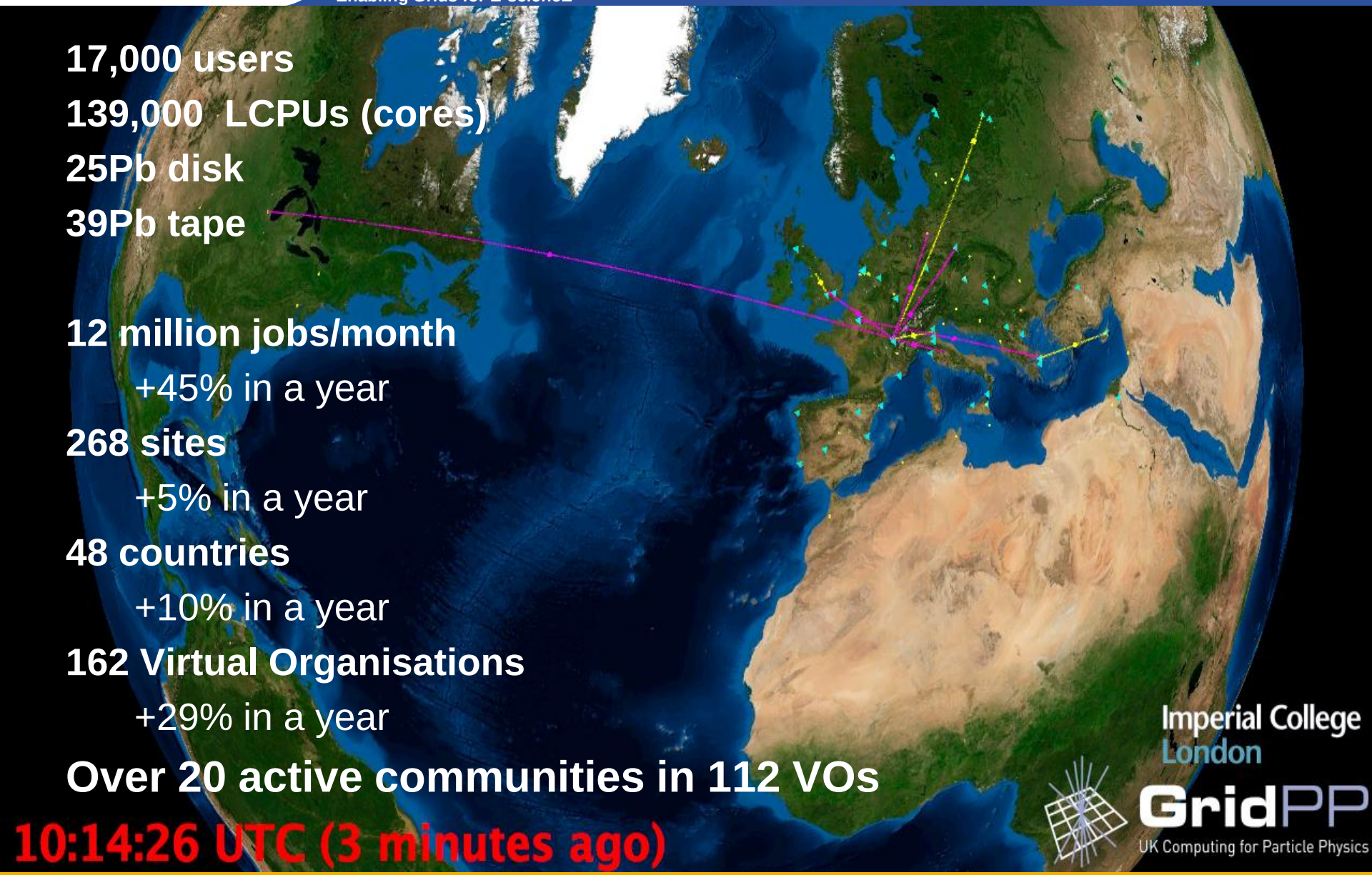
The Enterprise is the research community

European Grid Infrastructure

- **European Data Grid (EDG)**
  - Explore concepts in a testbed
- **Enabling Grid for E-sciencE (EGEE)**
  - Moving from prototype to production
- **European Grid Infrastructure (EGI)**
  - Routine usage of a sustainable e-infrastructure



EGEE has achieved a lot!



17,000 users

139,000 LCPUs (cores)

25Pb disk

39Pb tape

12 million jobs/month

+45% in a year

268 sites

+5% in a year

48 countries

+10% in a year

162 Virtual Organisations

+29% in a year

Over 20 active communities in 112 VOs

10:14:26 UTC (3 minutes ago)

Imperial College London

GridPP

UK Computing for Particle Physics

Lessons Learned

- **Supporting diverse communities is hard**
  - One middleware distribution (gLite) means compromises
  - Focusing on a single operating model provides tensions
- **Supporting a large operational infrastructure is costly**
  - Communication and coordination across 260+ sites
  - Running hardware: compute, storage, networking, ...
  - Running software: site, domain specific, ...
- **A production infrastructure does yield results**
  - Recent reconstruction events from the first LHC run
  - *In silico* drug discovery searches
  - Fusion simulations
  - .....

EGEE to EGI... what does it mean?

- **An opportunity!**
  - Draw a line under the experimentation in EDG & EGEE
  - Scope activities and structures so they are sustainable
- **A challenge!**
  - The technology landscape changes and we must change with it
  - Increasing diversity of application models and resources
    - Data Intensive Science is getting ever more intensive
    - Expand beyond core EGEE high throughput grids
      - *Integrate desktop and high performance grids*
    - Expand technologies in response to end-user & operational needs
      - *How do virtualisation and cloud computing change things?*
- **A business model!**
  - Add value where you can in providing a generic infrastructure
  - Provide an open extensible infrastructure for all

What will EGI initially focus on?

- Continue to provide a secure reliable generic infrastructure
  - Integrate resources based on gLite, UNICORE, ARC, Globus, ...
  - Leverage new technologies to provide more flexibility to users
- Support the user communities using the infrastructure
  - Assist and support the current EGEE communities
  - Engage with and support new structured communities
    - **e.g. ESFRI projects**
- Improve the efficiency of the infrastructure
  - The number of jobs, users & data continue to increase
  - Utilisation and effectiveness of the resources needs to match

Use new technologies to make middleware selection and operation a domain specific decision

European Strategy Forum on Research Infrastructures

- Roadmap updated in 2008
- Preparatory phase funding for most projects
- Big push in FP8 (2013 and beyond)?
- 44 projects covering:
  - Social Sciences and Humanities
  - Environmental Sciences
  - Energy
  - Biological and Medical Sciences

- Data Intensive Science
- National commitments in European context
- Global collaboration and shared access
- Long lifetime (10-20+ years)

g

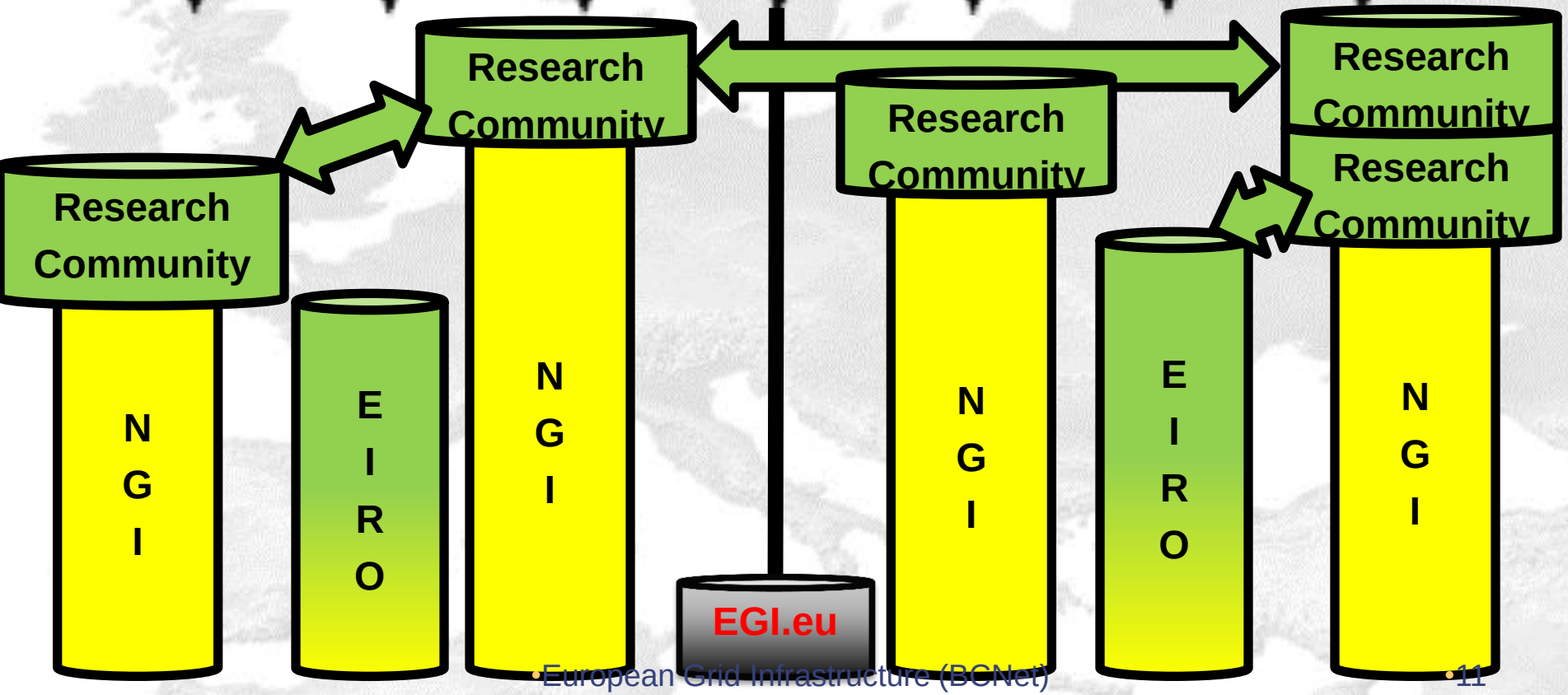
EGI means Innovation

- Deploy Technology Innovation
  - Distributed Computing continues to evolve
    - **Grids → Desktops → Virtualisation → Clouds →?**
- Enable Software Innovation
  - Provide reliable persistent technology platform
    - **Community tools built on the deployed technology**
- Support Research Innovation
  - Infrastructure for data intensive science
    - **Support for Virtual Research Communities (e.g. ESFRI)**

Virtual Research Communities

- Provide infrastructure to enable European research
  - Use the same infrastructure for national research
  - Link to other infrastructures to support European researchers
- Unit of representation within EGI
  - Discipline specific
  - Ideally aligned with an existing external structure
    - **e.g. EIRO, community, collaboration, project, ...**
  - Could include multiple VOs
  - Small international research groups
- Support through a federated helpdesk
  - Continued use of GGUS
  - Incorporate support units from:
    - **VRCs, Operations, Technology providers, Applications, ...**

EGI Collaboration



NGI: National Grid Initiative EIRO: European International Research Organisation

The EGI.eu Organisation

- Coordination for European DCI resources
  - Roadmap to integrate HTC, HPC, Data, Instruments, ...
  - Policy & services needed to run a production infrastructure
- EGI.eu governed and owned by its stakeholders
  - EGI Council votes proportional to fees
    - **Fees currently set proportional to national income**
  - Builds on resources from within its stakeholders
- Located in the Amsterdam Science Park
  - Distributed staff (~45) with a core (~50%) in Amsterdam
    - **Human coordination in Amsterdam**
    - **Technical coordination with a few partners across Europe**

EGI.eu's Services

- Integrated Infrastructure
  - Coordinates (not owns) the compute & storage resources
  - Resources owned by individual organisations
    - **They manage access for their user communities**
- Deploying Innovative Technology into Production
  - Software for secure authorised access to resources
    - **Liaison with external (to EGI) software providers**
    - **Integrated into the Unified Middleware Distribution (UMD)**
  - EGI defined and verified interfaces
    - **Compatible software must be deployed**
  - Interoperation within *your* NGI and internationally

EGI.eu's Services

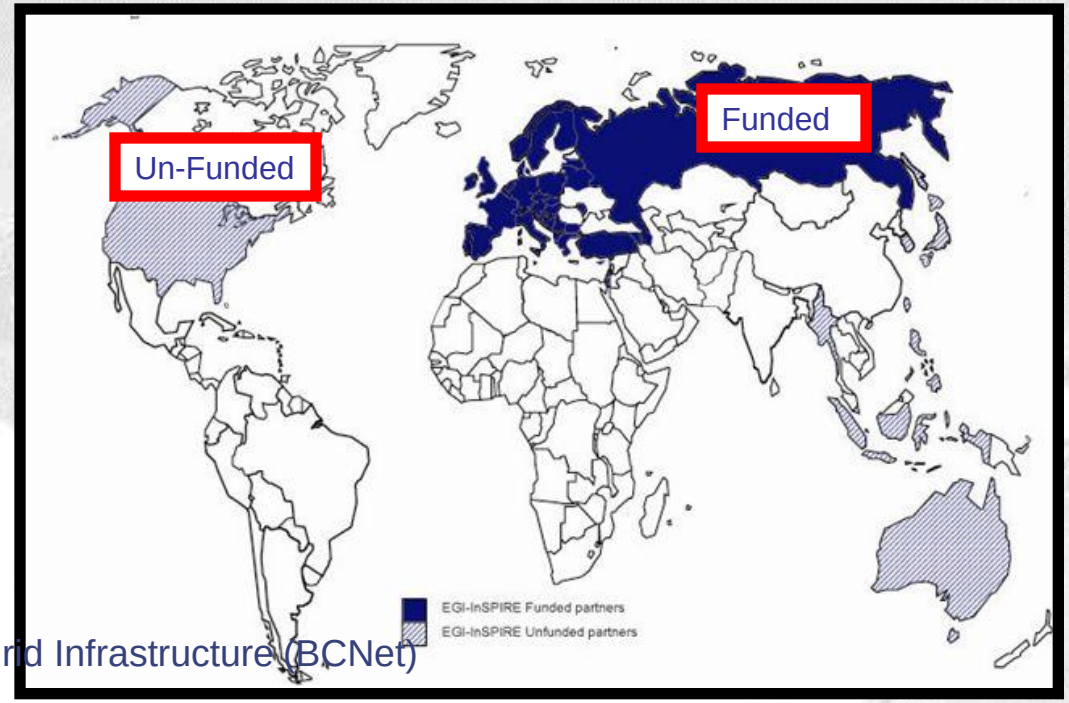
- User Community Support
  - From a single VO to a Virtual Research Community
  - Provide a federated Helpdesk linking:
    - **Discipline specific support (e.g. Bio Apps)**
    - **National infrastructure support (e.g. NGS)**
    - **Generic services within NGIs or VRCs (e.g. Training)**
  - Provide core services to support users
    - **Manage VOs, Application DB, Training DB**
  - Support for Heavy User Communities
- Dissemination
  - With NGIs, VRCs, and other projects
  - Two Annual meetings: Users & Technology
    - **EGI Technical Forum 14-17th September 2010 in Amsterdam**

The EGI-InSPIRE Project

Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

- A 4 year project with €25M EC contribution
  - Project cost €69M
  - Total Effort ~€330M
  - Effort: 9261

Project Partners (51)
- EGI.eu, 40 NGIs, 2 EIROs
- Asia Pacific (8 partners)



European Grid Infrastructure (BCNet)

Be a Neutral Infrastructure

- Consider IP network providers
  - Open to any traffic from many different communities
    - **Restrictions to protect other users**
  - Customised solutions within a generic framework
    - **Light paths on demand**
  - Standards drive integrated deployment
    - **Hardware and fibre from many different providers**
- And for sustainable E-Infrastructures?
  - Any application domain or middleware technology
  - A platform for domain specific innovation and use
  - Integration of any compliant compatible resources

Can we learn from others?

- Grids have benefited from commoditisation
  - Hardware: HTC & HPC affordable to all
  - Networking: GBs can be moved over WAN
  - Software: Open source software comes of age
- How will commodity virtualisation impact us?
  - For transactional models →
    - **Cloud Computing: A model based on compute not data**
  - For large distributed data-oriented models →
    - **The emergence of true 'function shipping'?**

Evolving Service Delivery Models

- Move towards an interoperable cloud infrastructure
  - Federated pan-European infrastructure
  - Use standards and the established AAAA mechanisms
- Provide a Data-Oriented Infrastructure as a Service
  - Use existing high performance data storage & transfers
  - Empower VRCs/VOs to source and run their own services
- Bring new research innovations into production
  - Federated cloud environments (i.e. VMs @ each site)
  - Experimenting with virtualised worker nodes in EGEE:
    - **e.g. INFN, BiG Grid, CERN, NGS, Dgrid, ...**

What does this evolution mean?

- EGI coordinates the core infrastructure
  - Assessing & certifying technology for deployment
  - Ensure integration of the core services in Europe
  - Operate & manage domain specific environments
    - **If required by that domain!**
- VOs now manage their own infrastructure
  - Decide what services are deployed where
  - Flexibility (& responsibility) to meet their own needs

Deregulate and open up the infrastructure
(Where it makes sense to do so!)

A long-term need for Standards

- Data Layer
  - Secure reliable data movement
  - Standardised access to data resources
- Virtualisation Layer
  - VMM across trust domains within agreed policies
  - Monitoring as important as lifecycle control
- Service Layer
  - The services that go into the vir
  - Avoid domain specific silos & pr

- Openness
- Consensus
- Balance
- Transparency

Sustainability

‘Europe as a hub for sustainable e-science and continuous service innovation’

- Reduce barriers for collaborative data intensive science
  - Integration with GEANT provides unique offering
  - Support to ESFRI projects and new communities
    - **Flexibility to run the services and software they need**
- Open global collaboration of e-infrastructures providers
  - Domain driven collaboration with other infrastructures
  - Open standardised interfaces to avoid vendor lock in
  - Add value where we can and outsource where we can’t

Summary

- EGEE:
  - Demonstrated a production e-infrastructure
- EGI:
  - Provide a sustainable production e-infrastructure
- EGI.eu is now a legal entity based in Amsterdam
  - Supported transition for 4 years through EGI-InSPIRE
- Contact: director@egi.eu

EGI Technical Forum

14-17th September 2010 in Amsterdam