European Grid Infrastructure: Enabling the Global Research Community

Steven Newhouse  
Technical Director, EGEE-III  
Director, EGI-InSPIRE  
Interim Director EGI.eu
What is e-Infrastructure?

- **Resources linked by high speed networks**
  - Compute, Storage, Instruments, ...
- **Controlled access to shared resources**
  - Authentication, Authorisation, Accounting, ...
- **Dependable services for others to use**
  - Driven by availability and reliability metrics
- **Services that are there for the long-term**
  - Supporting experiments lasting decades
Enabling Grids for E-sciencE

European Grid Infrastructure

• European Data Grid
  - EGEE-I
  - SEE-GRID
  - Baltic Grid I
  - Nordic DataGrid Facility
  - EGI_DS

• European Grid for E-sciencE (EGEE)
  - EGEE-II
  - SEE-GRID-2
  - Baltic Grid II
  - EGI_InSPIRE

• European Grid Infrastructure (EGI)
  - Routine usage of a sustainable e-infrastructure

Outreach to Asia Pacific

European Grid Infrastructure (ISGC 2010)
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)
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
    ▪ Encompassing desktop and high performance grids
    ▪ How do virtualisation and cloud computing models 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?

• 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 ESFRI projects to support their requirements

• Improve the efficiency of the infrastructure
  – Jobs, users & data continue to increase
  – Utilisation and effectiveness of the resources needs to match

Explore new technologies to move towards middleware selection and operation being a domain specific decision
EGI means Innovation

• **Deploy Technology Innovation**
  – Distributed Computing continues to evolve
    • Grids → Desktops → Virtualisation → Clouds →?

• **Enable Software Innovation**
  – Provide reliable persistent technology platform
    • Today: Tools built on gLite/UNICORE/ARC
    • RESPECT: Recommended External Software
    • for Egee EGI CommuniTies

• **Support Research Innovation**
  – Infrastructure for data intensive science
    • Support for international research (e.g. ESFRI)
European Strategy Forum on Research Infrastructures

• Updated roadmap in 2008 with 44 projects
• Preparatory phase funding for most projects
• Big push in FP8 (2013 and beyond)?
Technology Innovation

• Will come from outside EGI
  – EGI runs Distributed Computing technologies in production

• Partnership with technology projects
  – EMI (European Middleware Infrastructure)
    • Continued evolution of gLite, ARC, UNICORE
  – IGE (Initiative for Globus in Europe)
    • Support and customisation for European Globus users
  – StratusLab
    • Use RESERVOIR to support gLite environments
Collaborative Development

Technology innovations taking place outside the production infrastructure

Innovating technology being deployed within the production infrastructure

Prototype

Usage

Deployment

Release

Evaluate

Requirements

Implementation

Feedback

EC & National Research Projects

NGIs & EIROs

Collaboration

(ISGC 2010)
Software Innovation

• Will come from outside the infrastructure provider
  – EGI is a neutral platform for applications

• EGI cannot support all services in its core
  – Every community needs something different

• Support & foster innovation within different ‘sectors’
  – High Throughput Computing
    • gLite, ARC, ...
  – High Performance Computing
    • UNICORE, ...
  – Digital Libraries
    • gCube from D4Science
Research Innovation

• An infrastructure to support European Researchers
  – Within the EU27
  – Geographical Europe
  – Interoperability worldwide for collaboration
• Work with Virtual Research Communities
  – Groupings of aligned Virtual Organisations
  – Provide community specific:
    • Support, training, consultancy, requirements etc.
Collaboration
The EGI.eu Organisation

• Coordination for European Grid resources
  – Roadmap to integrate HTC, HPC, Data, Instruments, ...
  – Policy & services needed to run a grid
• Governance & ownership by its stakeholders
  – EGI Council votes proportional to national income
  – EGI Council fees proportional to votes
  – Sustainable small coordinating organisation (EGI.eu)
  – Builds on resources from within its stakeholders
• Location selected to be Amsterdam
  – Distributed staff with a core (~50%) in Amsterdam
    • Human coordination in Amsterdam
    • Technical coordination with a few partners across Europe
  – Approximately 45 staff, €3.5M/year
EGI.eu’s Services

• User Community Support
  – Scale up 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
EGL.eu’s Services

• Integrated Infrastructure
  – EGL.eu does not own or control the compute & storage resources
  – Resources owned by individual organisations
    • They manage access for their user communities

• Deployed Technology
  – Software for secure authorised access to resources
    • Liaison with external (to EGL) software providers
    • Integrated into the Unified Middleware Distribution (UMD)
  – EGL defined and verified interfaces
    • Compatible software must be deployed
  – Interoperation within your country and internationally
The EGI-InSPIRE Project

Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

• A 4 year project with €25M EC contribution
  – Project cost €70M
  – Total European Grid Effort ~€330M
  – Global Scope: 41 partners
    • EGI.eu, 37 NGIs, 2 EIROs, APGI (~11 partners, 8 countries)

• Project effort:
  – 8138PM mixed EC/NGI/EGI.eu funding model
  – 1100PM unfunded Asia Pacific Grid Initiative

• Provide the framework to deploy innovation in DCIs
  – Research: Enable Data Intensive Science worldwide
  – Technology: Deploy new approaches: Grids, Desktops, Clouds, ...
  – Software: Bring external components into production deployment

European Grid Infrastructure (ISGC 2010)
Project Partners (41)
- EGI.eu, 37 NGIs, 2 EIROs
- APGI (represents ~11 partners, 8 countries)
- OSG: In spirit but not a formal partner!
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

• European Grid Infrastructure (ISGC 2010)
Supporting Multiple Communities

Use Cloud technology to provide a minimal set of ‘Core Site Services’
Supporting Multiple Communities

Core Site Infrastructure Services: European Cloud Infrastructure
A Cloud for Data Intensive Science?

Coordination by EGI.eu
Technology assessment, Integrated Operations & User Support

VMM:
Virtual Machine Managers
European Cloud Infrastructure

• Underpinned by an interoperable cloud infrastructure
  – Sites deploy the VM management technology they want
  – Securely integrated into a reliable infrastructure
  – Accessible to authenticated, authorized & accounted for users

• Provide a Data-Oriented Infrastructure as a Service
  – Leverage existing high performance data storage & transfers
  – Application domains (VOs) source and run their own services
  – VO Managers deploy & run these services on the infrastructure

• Core resources from research centres within Europe
  – Able to ‘burst out’ to other countries and commercial providers
Enable new Innovation

• Secure platform for launching VMs across Europe
• Production testbed for new site VM managers
• Provide domain specific environments on demand
  – Experimenting with virtualised worker nodes in EGEE:
    • e.g. INFN, BiG Grid, CERN, ...
  – Expand to provide other site services on demand
Sustainability

- **Business model of providing infrastructure services**
  - The core needed to integrate different resource centres
- **Adapt the infrastructure to the expanding user base**
  - Provide a platform for launching VMs
- **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**
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