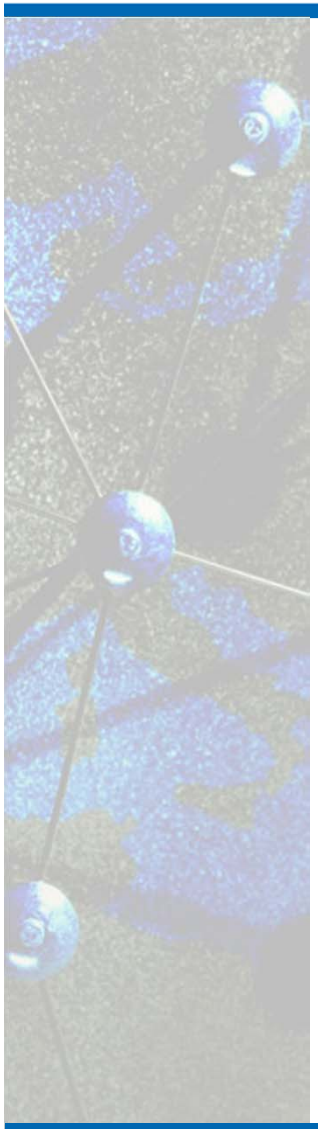


EGI-InSPIRE Project Presentation

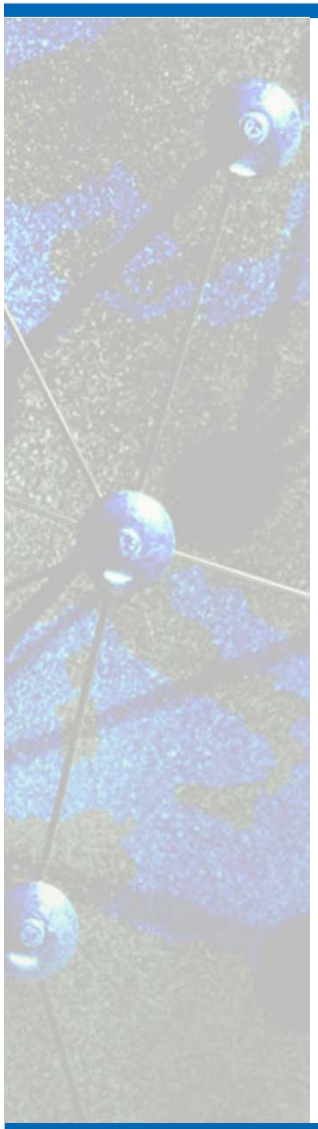
Steven Newhouse

Project Director, EGI.eu

e-infrastructure

Abbreviations

- EGI: European Grid Infrastructure
- EGI.eu: European Grid Initiative organisation
- EIRO: European International Research Organisation
- ESFRI: European Strategy Forum on Research Infrastructures
- HUC: Heavy User Community
- NGI: National Grid Infrastructure/Initiative
- RP: Resource infrastructure Provider
- SSC: Specialised Support Centre
- UMD: Unified Middleware Distribution
- VO: Virtual Organisation
- VRC: Virtual Research Community



Why build a European Grid 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

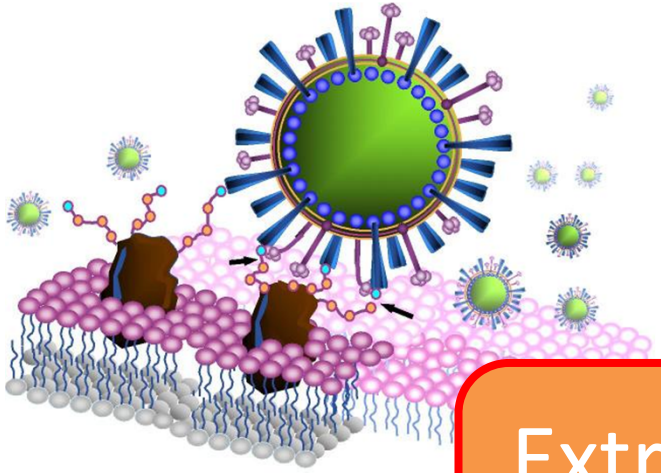
The Enterprise is the European Research Area

EGI provides a service infrastructure that exposes and helps coordinate a resource infrastructure

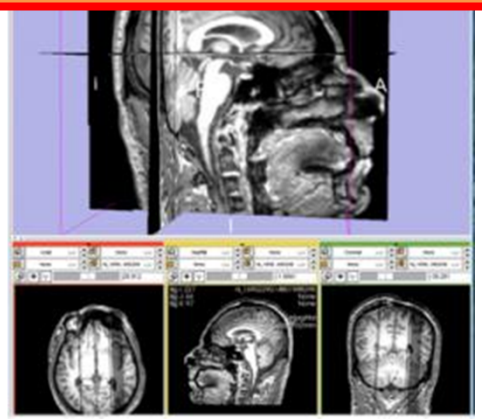
What is a Grid?

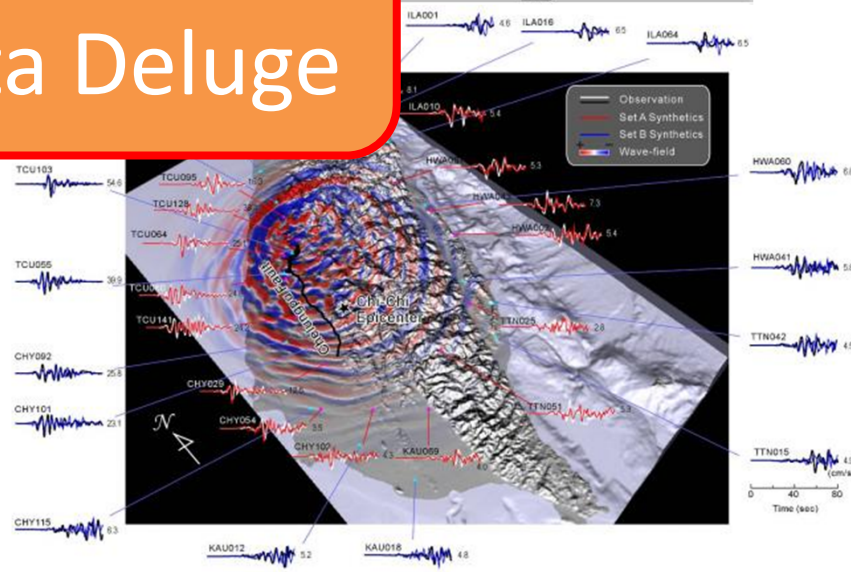
- A grid consists of distributed resources controlled by separate organisations that be systematically used securely by users external to that organisation
- Resources can include:
  - Commodity or HPC clusters
  - Disk or tape storage
  - Instruments
  - Data Archives or Digital Libraries

C21: Digital Research

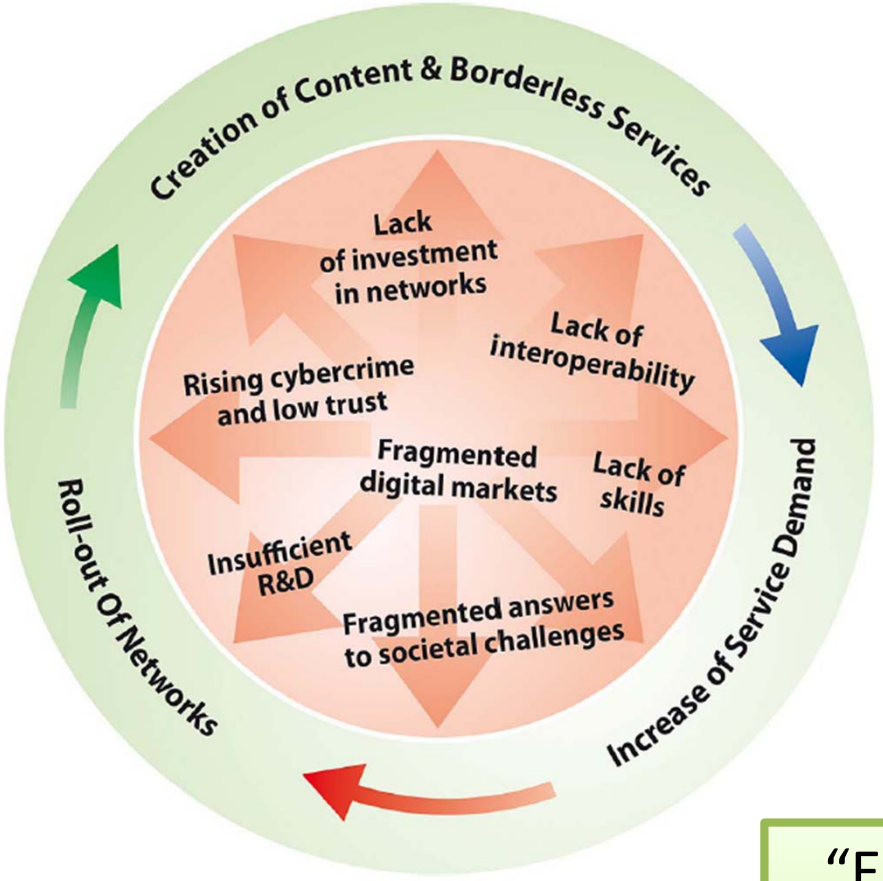


Extracting Knowledge from the Data Deluge





Digital Agenda for Europe

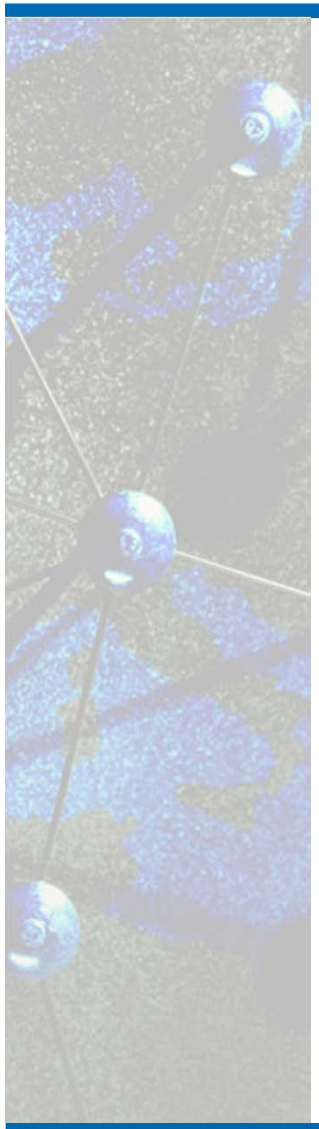


Digital Agenda for Europe

- Borderless Services
- Interoperability
- Supporting Innovation

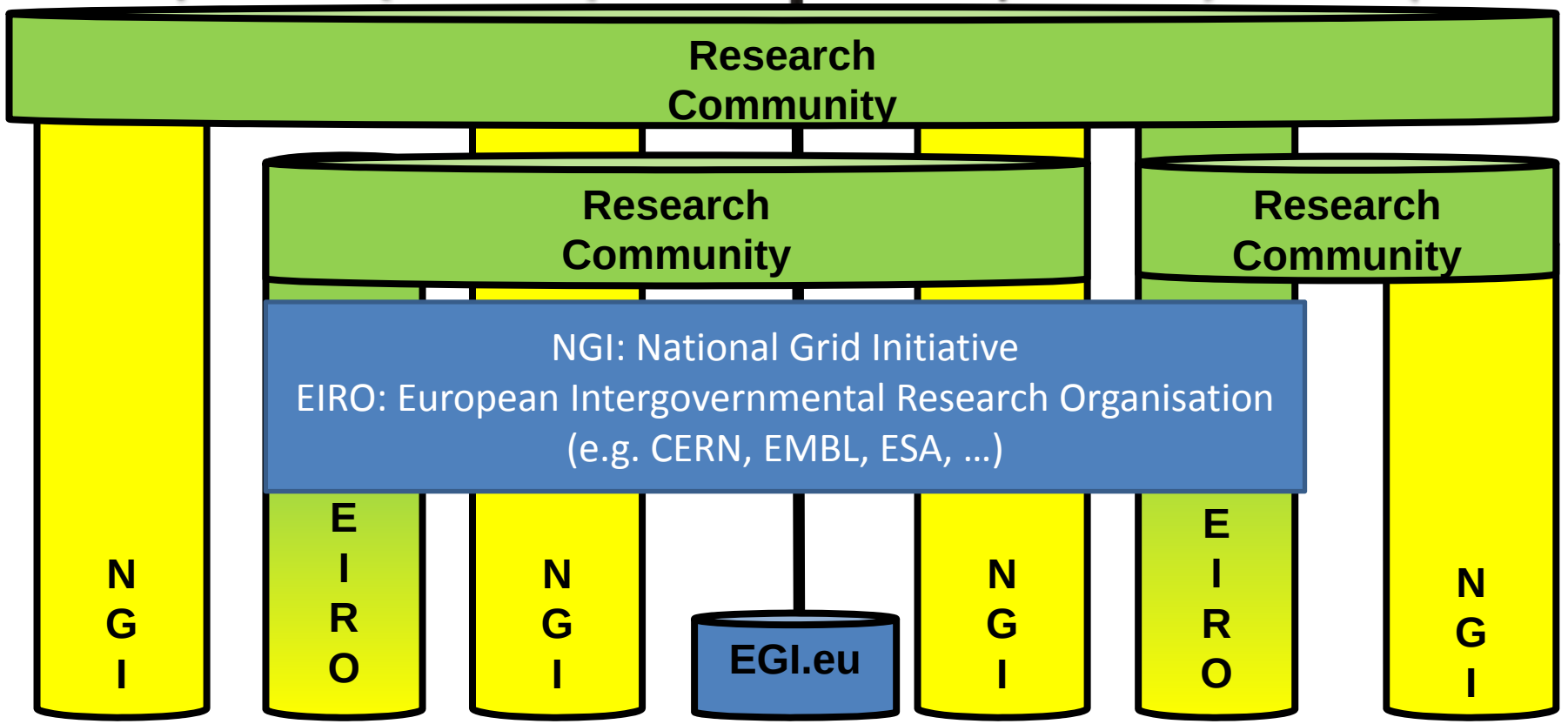
remove barriers to the free movement of knowledge

"Europe should also build its innovative advantage in key areas through reinforced e-Infrastructures (i.e. GEANT & EGI)"

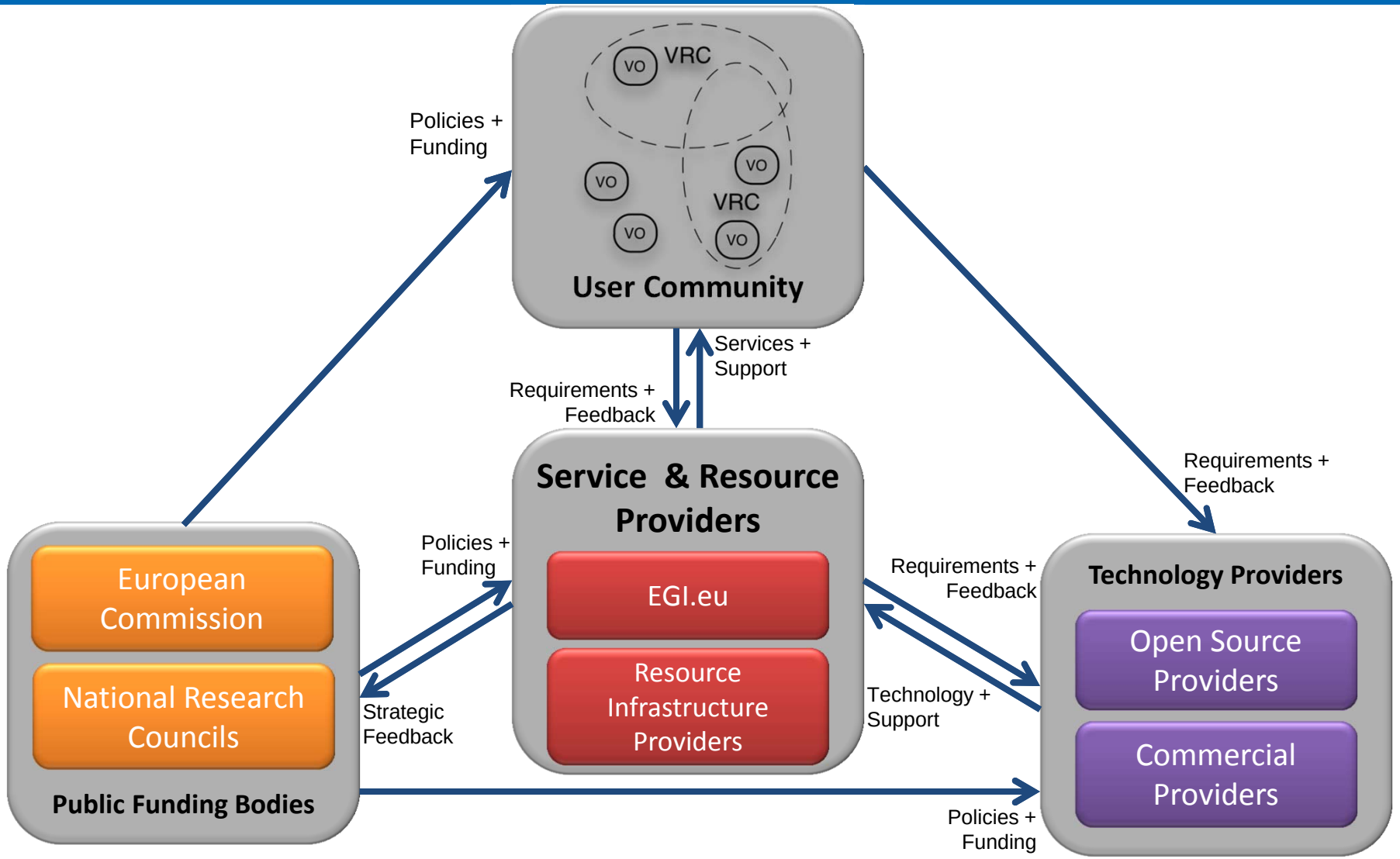


The EGI Model

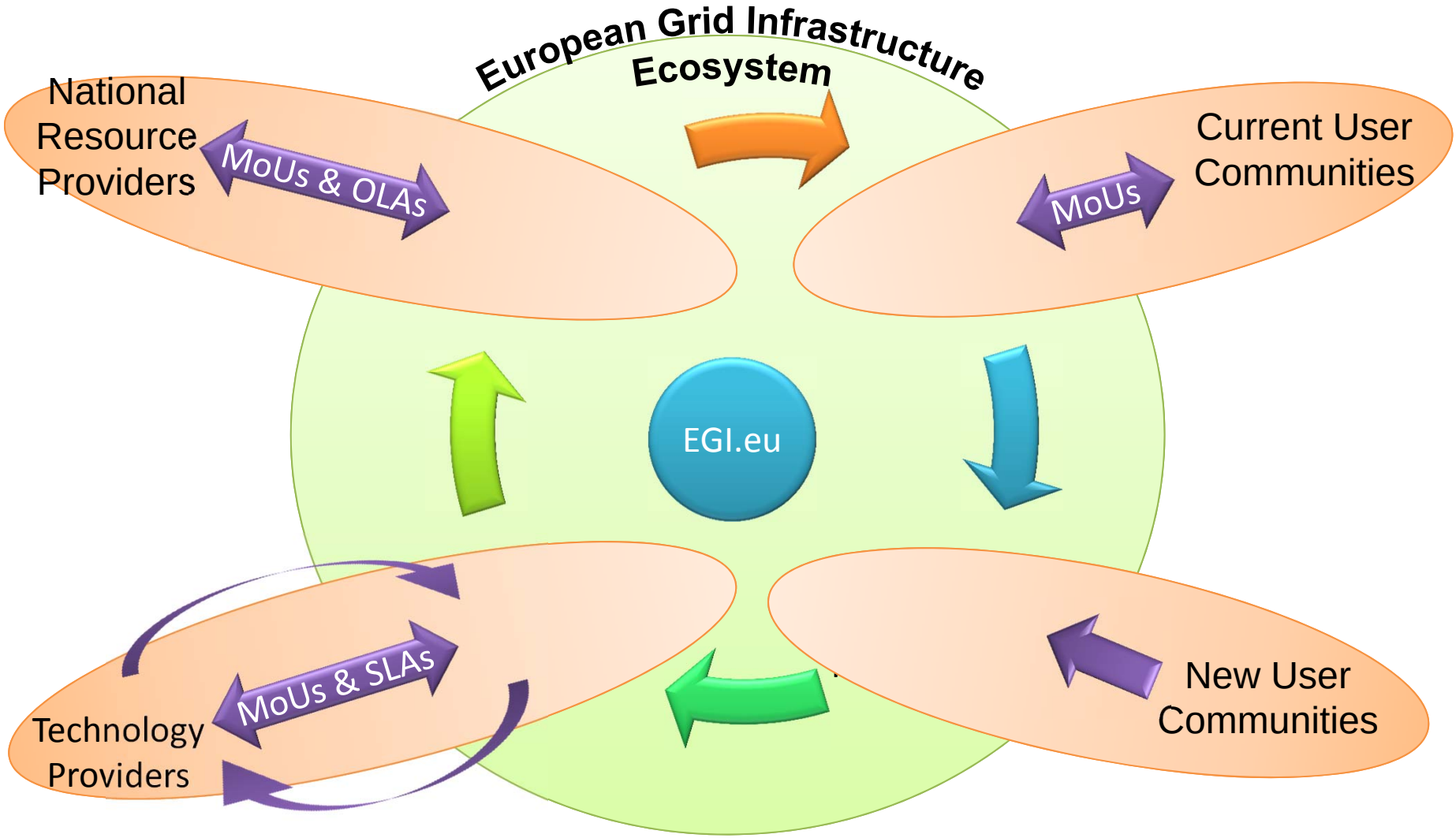
EGI Collaboration



EGI Ecosystem

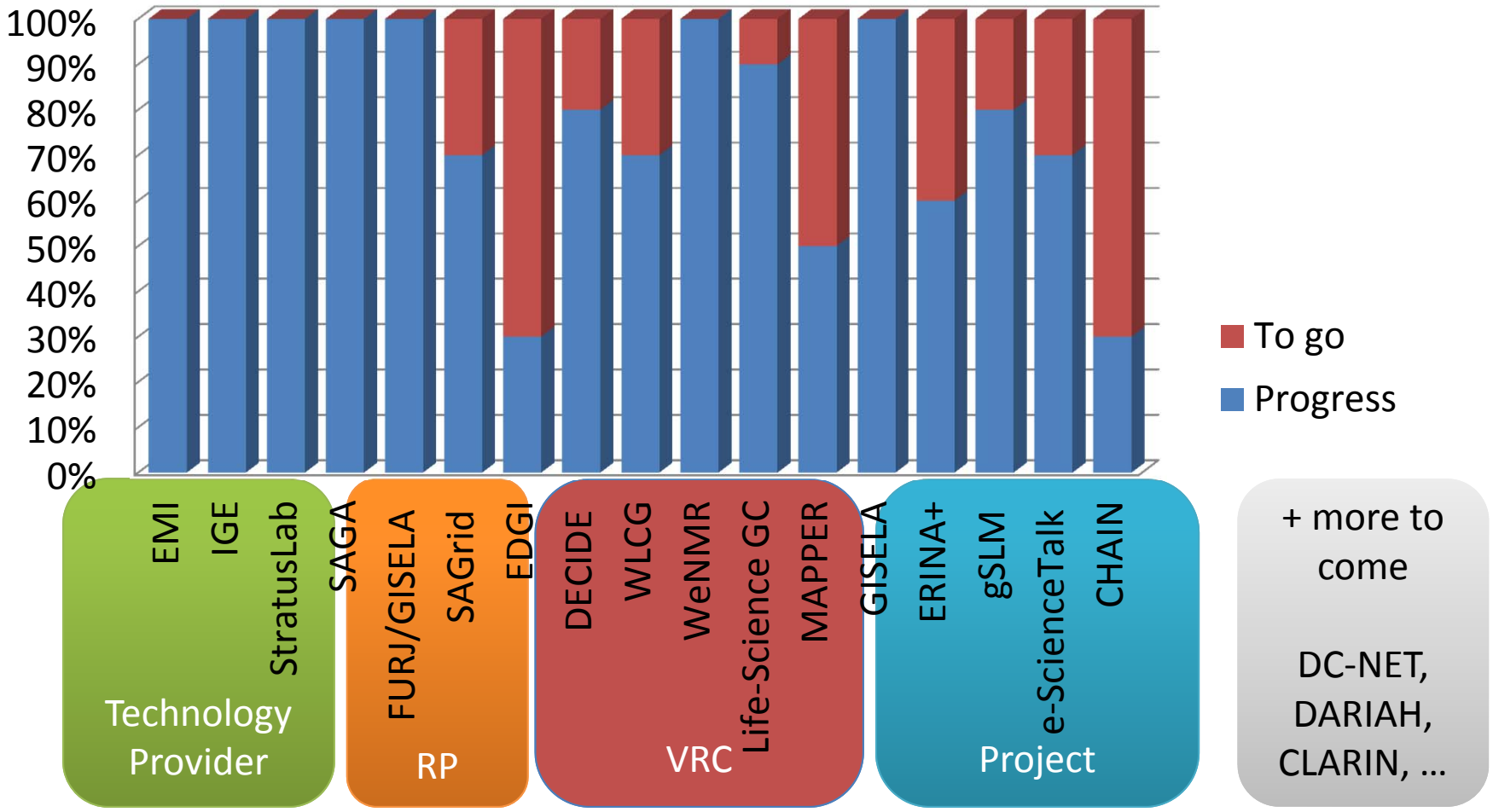


A Virtuous Service Cycle



MoUs Structure Ecosystem

http://www.egi.eu/collaboration/



EGI.eu

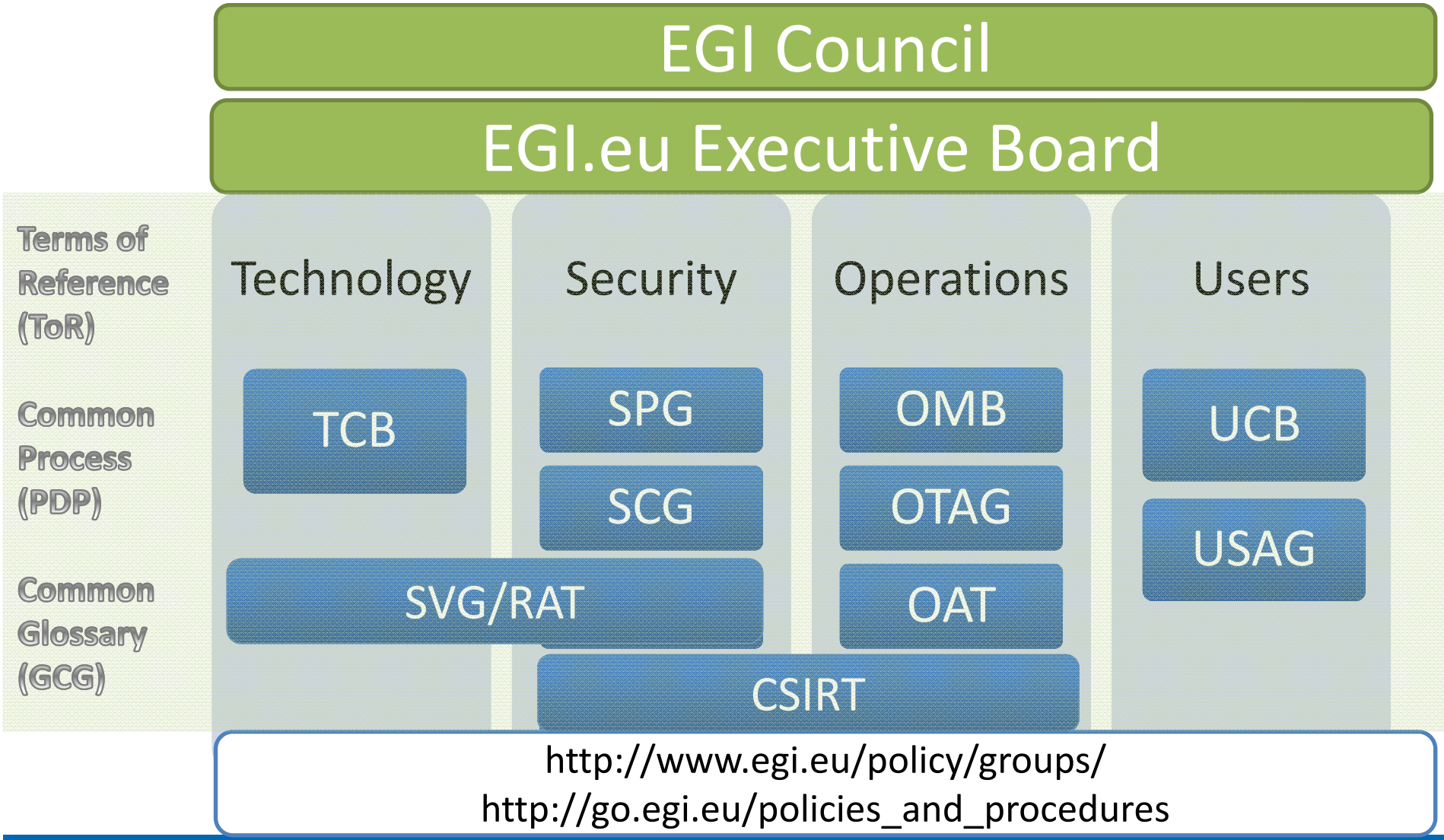
- Coordination for European Grid resources
  - Established February 8th 2010
  - Central policy & services needed to run a grid
  - Sustainable small coordinating organisation
- Based in Amsterdam
  - Coordinating core (~20 people) in Amsterdam
  - Technical services from partners (~20 people)

EGI and EGI.eu: Supported by the EGI-InSPIRE project

EGI.eu Governance

- EGI.eu established as non-profit foundation
- Governance & ownership by its participants
  - Participants:
    - European NGIs
  - Associated participants:
    - Organisations aligned with EGI.eu's objectives
- EGI Council contains all participants
  - Votes linked to fees

EGI Technical Governance



EGI-InSPIRE Project

Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

A 4 year project with €25M EC contribution

- Project cost €72M
- Total Effort ~€330M
- Effort: 9261PMs

Project Partners (50)

EGI.eu, 38 NGIs, 2 EIROs

Asia Pacific (9 partners)

Project Objectives

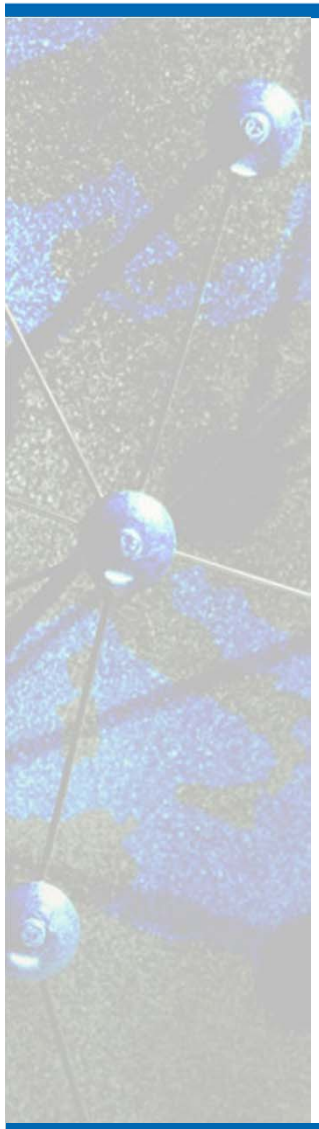
- A sustainable production infrastructure
  - Resource providers in Europe and worldwide
  - With new technologies as they mature
- Support structured international research
  - Sustain current domain specific services
  - Attract new user communities (e.g. ESFRI)

Project Activities

- NA1: Project & Consortium Management
  - Project Office and Quality Assurance
- NA2: External Relations
  - Policy Development and Dissemination
  - Community Building Events
- NA3: User Community Coordination
  - EGI.eu and NGI support teams
  - Supporting Technical Services for Virtual Research Communities
- JRA1: Support for Operational Tools
  - Maintenance and Development
  - Support for new resources and their accounting

Project Activities

- SA1: Operation of the production infrastructure
  - Infrastructure oversight and quality control
  - Operational security
  - Operational Tools, monitoring & accounting
  - Helpdesk & Support teams (NGI & centrally)
  - Validation and integration of new technology
- SA2: Provisioning the Software Infrastructure
  - Definition of software coming from external projects
  - Validation of delivered software
  - Software repository and support tools
- SA3: Support for Heavy User Communities
  - Services & tools for all users of the infrastructure
  - Domain specific support for current heavy users



What does EGI do?

User Support & Services

- **Support User Communities**
  - Researchers in International Collaborations
  - National Research Collaborations through the NGI
  - Scale up from the single VO to a community
- **Provide core services to support users**
  - Manage VOs, AppDB, Training Services
- **Support teams**
  - EGI.eu User Community Support Team
  - NGI User Support Teams
  - NGI Operations Teams
  - Experts within user communities or projects

A Virtuous User Cycle

Aka: "the chicken and egg conundrum...

Feedback through VRCs in the User Community Board

User Services Advisory Group to drive detailed design

Discover

Where is the community?

- VRCs
- Mailing lists
- Workshops
- Forums
- Blogs
- Projects
- Sharing stories
- Collaborating

Design

How can I contribute?

- Applications
- Data collections
- Requirements
- Proposals
- Projects
- Success stories

Deliver

Integrated Services

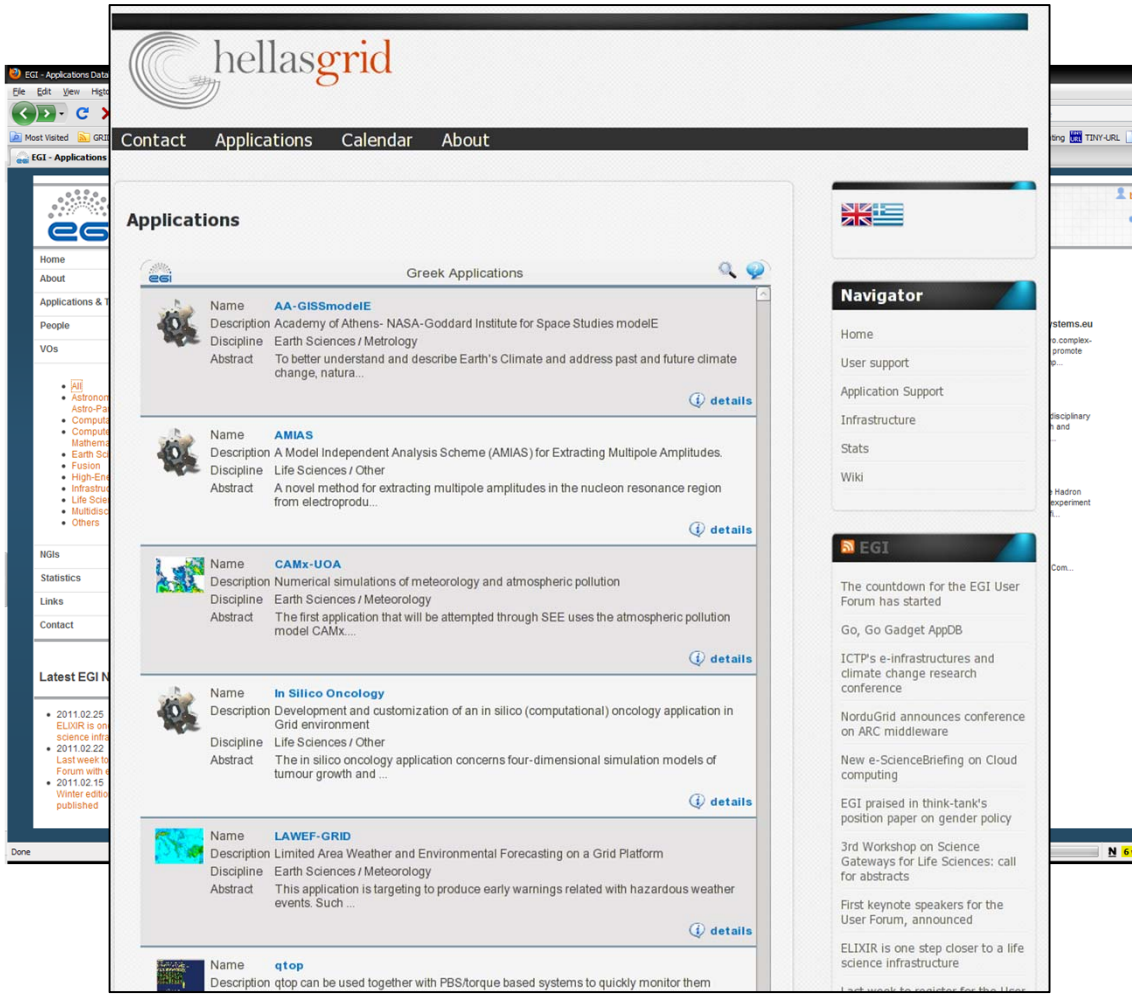
- Human
- Technical
- Infrastructure

How do I use these resources?

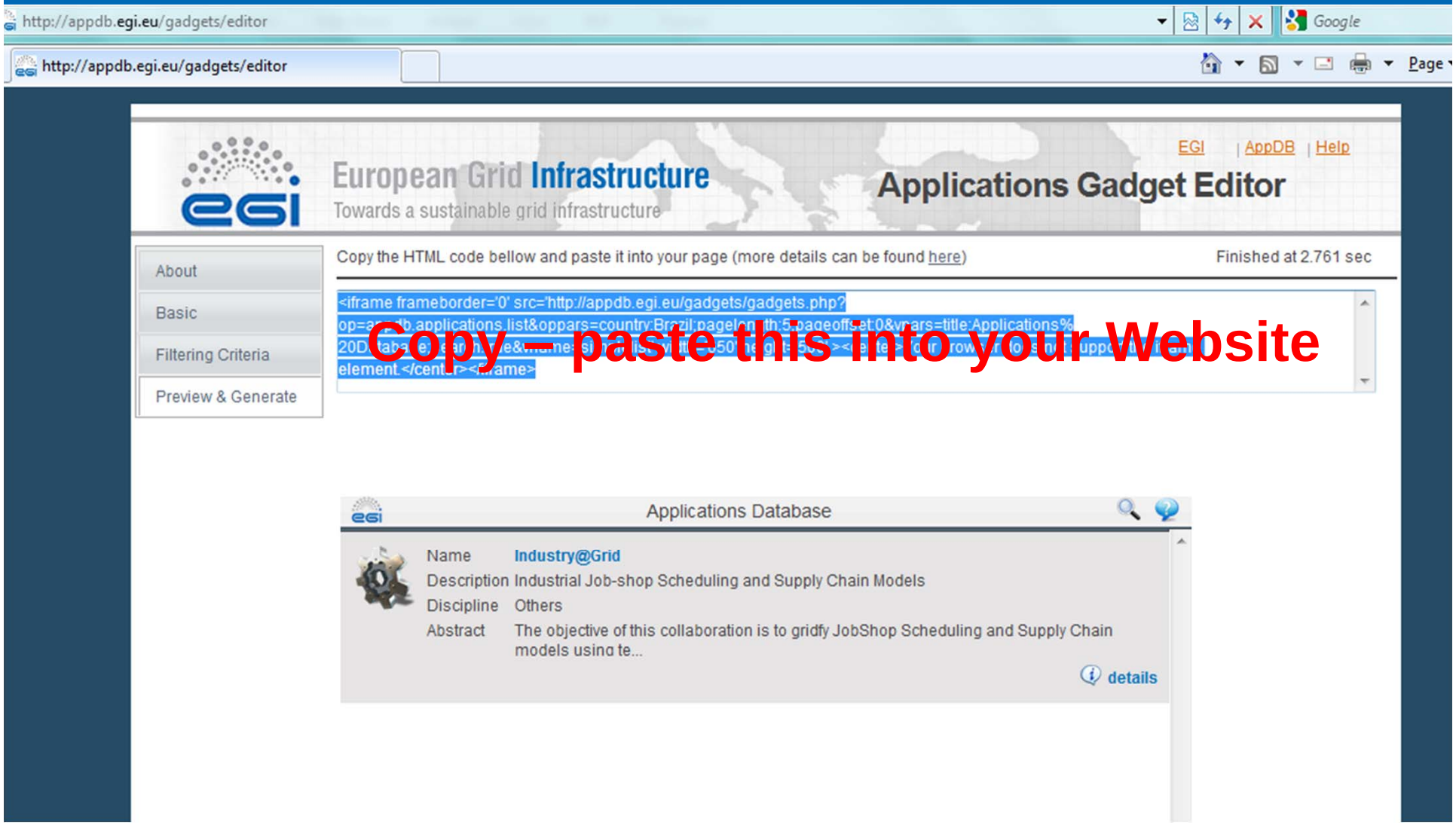
- Attend training courses
- Utilise training material
- Access data
- Run applications on the grid

EGI Application Database

- To give recognition to reusable applications
- To give recognition for application developers
- How to get involved
  - Register applications
  - Reuse applications
  - Integrate AppDB through its gadget into any Webpage!
- http://appdb.egi.eu

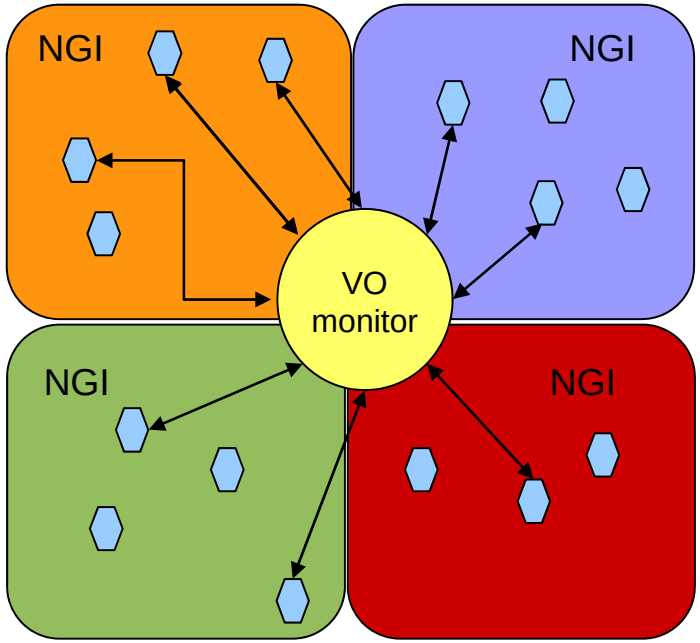


The AppDB gadget



Services for VOs

- Activities
  - Consultancy and helpdesk for VO managers
  - Evaluation of VO management, monitor and accounting tools
  - Provision of VO support software for VRCs
- VO-specific monitoring
  - Monitor only those sites that support you
  - Create and plug-in VO-specific probes
- How to get involved?
  - Request support
  - Prepare and share reviews of VO tools
  - Offer local solutions for VOs through the group



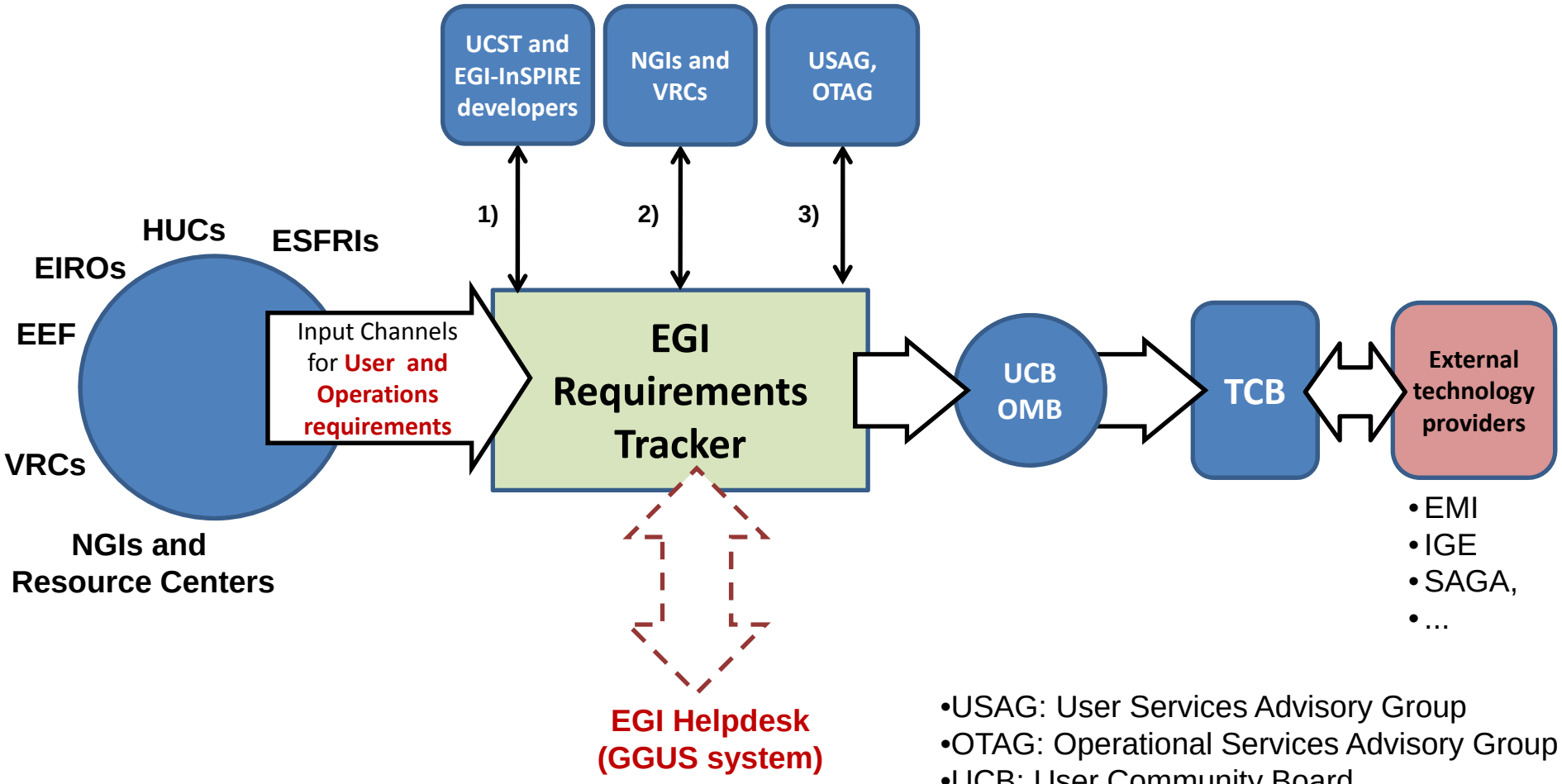
https://wiki.egi.eu/wiki/VO_Services

Training Marketplace

- Integration of…
  - Training event calendar
  - Digital library (of training materials)
  - Service offering and requesting form
  - Plus: community contributed evaluation and reviews
- How to get involved?
  - Register training events
  - Use, reuse and share training materials
  - Offer training-related services (infrastructure, VO, CA, ...)
  - Contribute training-requirements
  - Delegate representative to EGI Training Working Group

http://training.egi.eu

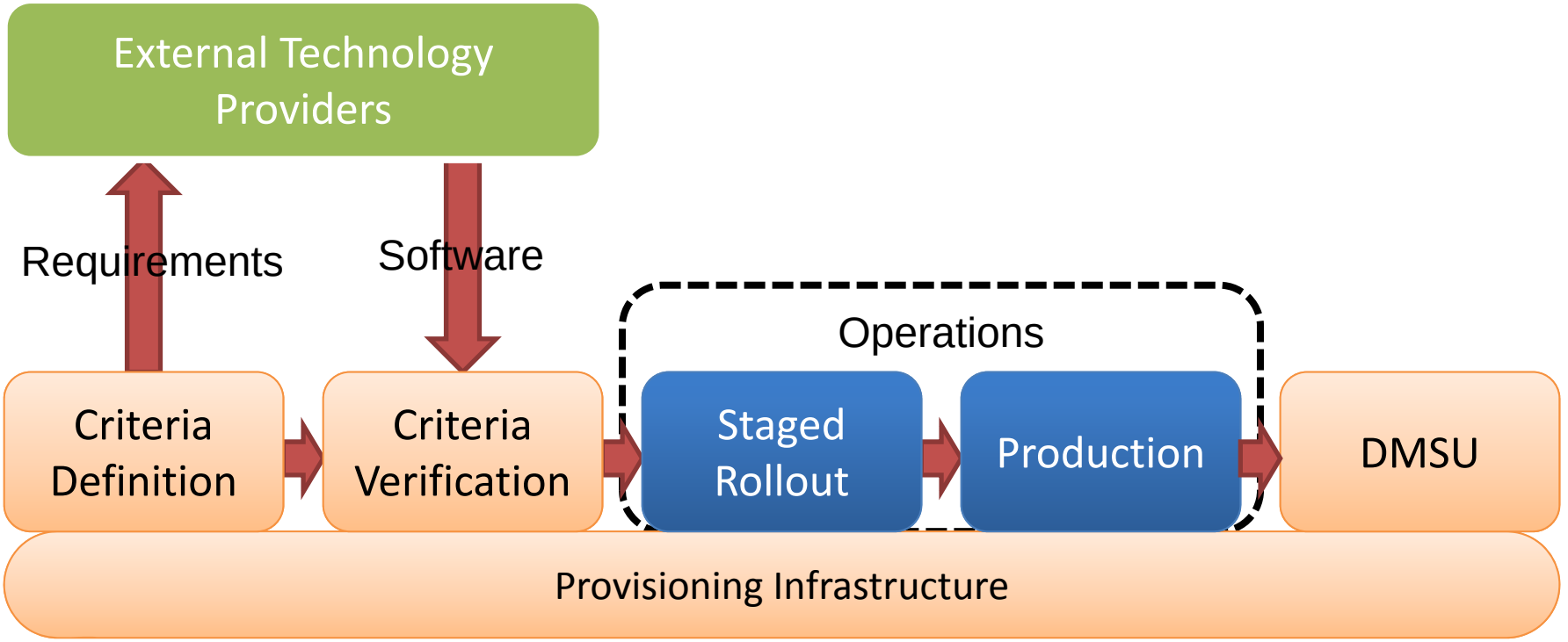
Requirements Lifecycle



•USAG: User Services Advisory Group
•OTAG: Operational Services Advisory Group
•UCB: User Community Board
•TCB: Technology Coordination Board

Software Provisioning

Integrated processes for efficient software provisioning



Technology Environment

- EGI Technology Roadmap
  - Technical environment & its evolution
  - Mainstream open-source components
  - Open to commercial components
- Unified Middleware Distribution (UMD)
  - Components from within the EGI community
    - External technology provider: EMI, IGE, SAGA, …
    - Details contained in the UMD Roadmap
  - To meet the unique needs of EGI users

Provisioning Workflow

- Infrastructure for verifying UMD components
  - Technology Provider triggers the process
- Criteria Verification organises and executes the verification workplan
- StagedRollout exposes verified software in the production infrastructure
- UMD release is assembled and published

Quality Criteria Definition

- Comprehensive library of Quality Criteria
- Defines EGI's minimum acceptance level of any delivered software
- Library with defined lifetime, and version
- Liaising with Technology Providers for
  - Proactive Quality Assurance
  - Quality Criteria Review and improvement
  - Timely preparation for software changes

Criteria Verification

- Formal verification of software using the Quality Criteria library valid at that time
- Publishes verification reports
- Verification reports provides feedback to
  - Criteria definition
  - Technology Providers
  - StagedRollout
  - Deployed Middleware Support Unit (DMSU)
  - UMD Release Notes

Provisioning Infrastructure

- UMD Repository (http://repository.egi.eu/)
  - Frontend publishes release notes etc.
  - Temporary internal repositories for:
    - Verification & StagedRollout
  - Composite repository for public UMD updates
- Process management infrastructure
  - GGUS → External Technology Providers
  - RT → Internal Verification and StagedRollout

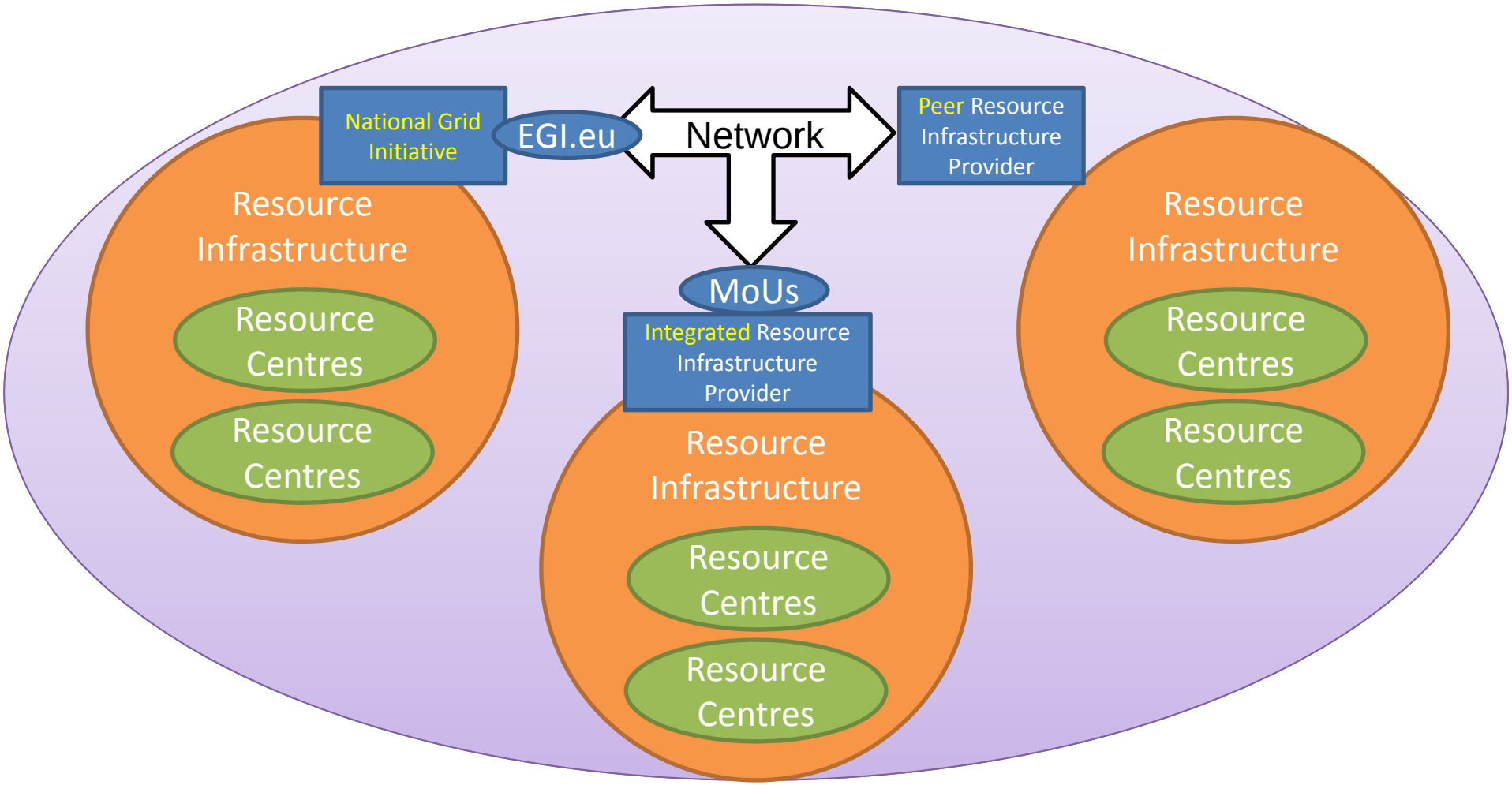
Deployed Middleware Support Unit

- Second level support unit for middleware
- Resolves issues with documentation, configuration
- Provides workarounds for known issues
- Develops and publishes best practices for popular components
- Provides recommendations for Technology Providers via the TCB

Services for HUCs

- Supporting major production computing
  - At an unprecedented scale – both quantitatively and qualitatively
- Delivering common solutions across multiple communities
  - Identifying areas for future work
- Broadening the use of grid technology and HUC services to related projects within the HUC domain
  - Especially unfunded Life Science & Earth Science projects
- Developing a S.W.A.T. analysis of each HUC
  - Focus steps on the road to sustainability

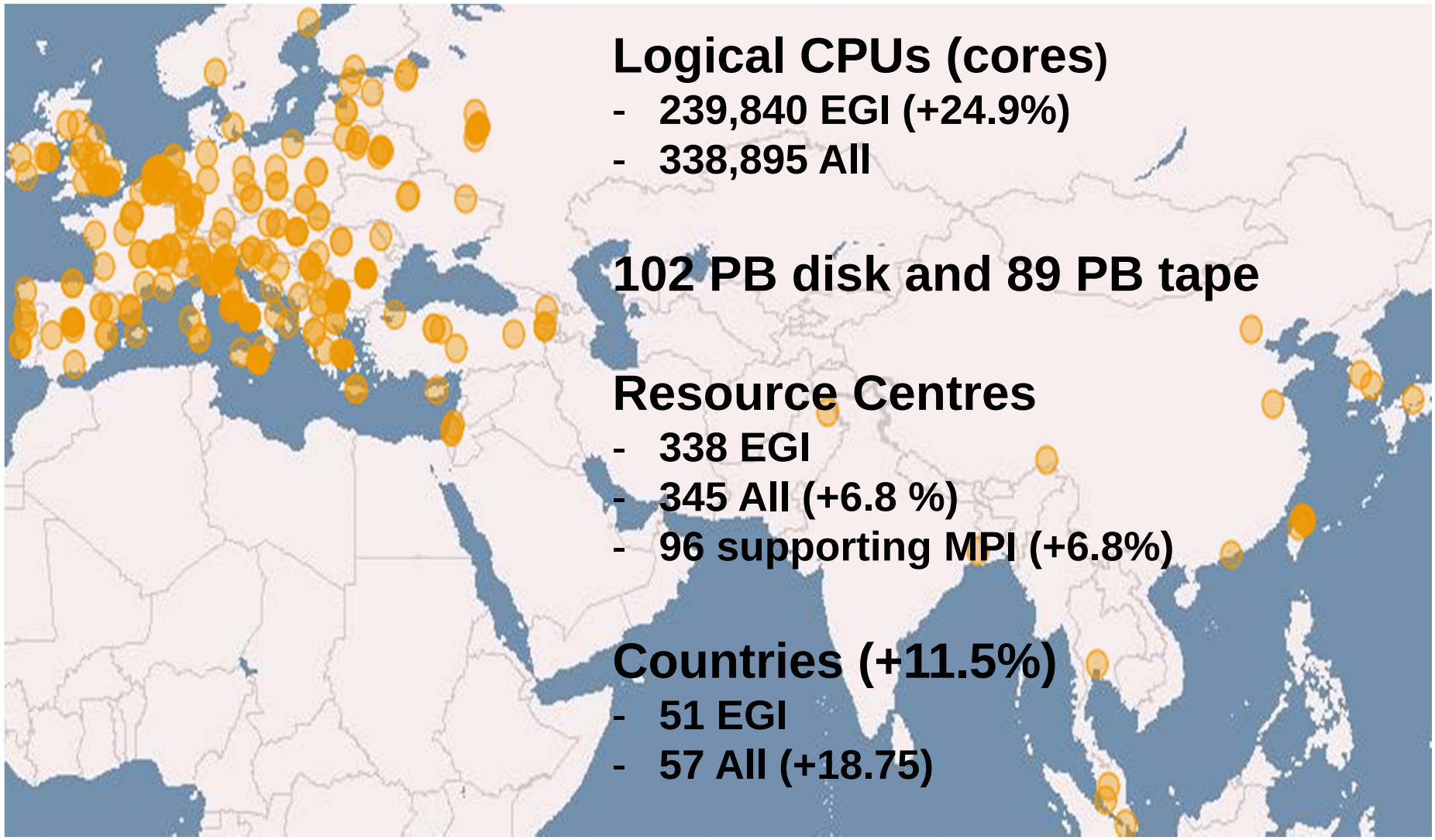
EGI Resource Infrastructure



Integrated: infrastructure operated by a non-EGI-InSPIRE partner but relying on EGI operational services (MoU)
Peer: infrastructure accessible to EGI users, but relying on own operational services

European Grid Infrastructure

(April 2011 and yearly increase)



Logical CPUs (cores)

- 239,840 EGI (+24.9%)
- 338,895 All

102 PB disk and 89 PB tape

Resource Centres

- 338 EGI
- 345 All (+6.8 %)
- 96 supporting MPI (+6.8%)

Countries (+11.5%)

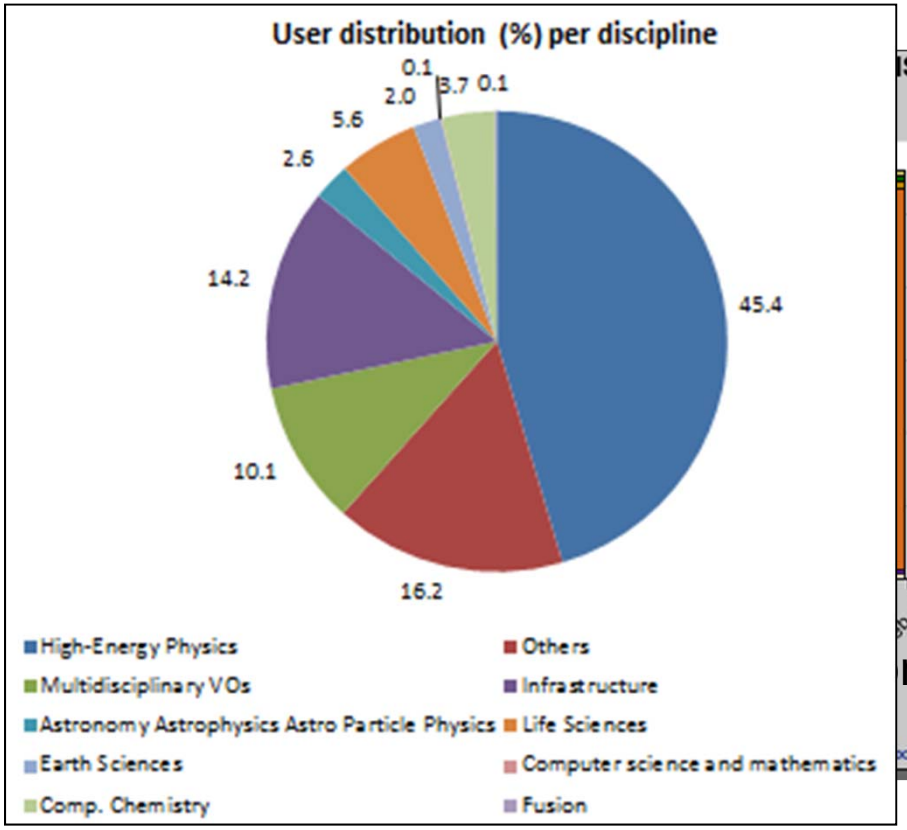
- 51 EGI
- 57 All (+18.75)

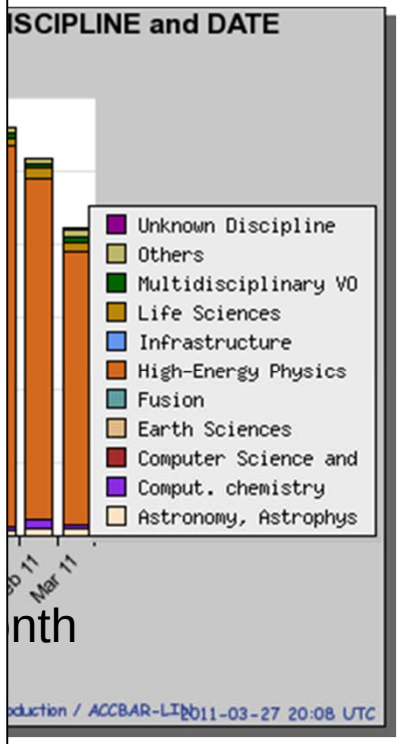
EGI Usage
(April 2011)

11226 End-users (-7.7%)
219 VOs (+17.7%)
~30 active VOs: constant

User Communities

- Archeology
- Astronomy
- Astrophysics
- Civil Protection
- Comp. Chemistry
- Earth Sciences
- Finance
- Fusion
- Geophysics
- High Energy Physics
- Life Sciences
- Multimedia
- Material Sciences
- ...





Average usage 2010-2011 vs 2009-2010

- 26.6M jobs/month, 873,400 jobs/day (+420%)
- 74.6M CPU wall clock hours/month (+86.5%)
- 551M HEP-SPEC06 CPU wall clock hours/month (+99.4%)

EGI Resource Providers and Operations Centres 1/2

- **22 National Operations Centres**
  - Armenia, Belarus, Bosnia and Herzegovina, Bulgaria, Cyprus, Croatia, Czech Republic, France, FYR of Macedonia, Germany, Georgia, Greece, Hungary, Israel, Italy, Montenegro, Poland, Romania, Serbia, Slovakia, Slovenia and Turkey
- **5 Federated Operations Centres** (16 NGIs)
  - IberGrid (Portugal and Spain)
  - The Netherlands Federation (Belgium and The Netherlands)
  - Russian Federation (Russia and Ukraine)
  - NDGF Federation (Austria, Denmark, Estonia, Finland, Latvia, Lithuania, Norway and Sweden)
  - United Kingdom/Ireland Federation

EGI Resource Providers and Operations Centres 2/2

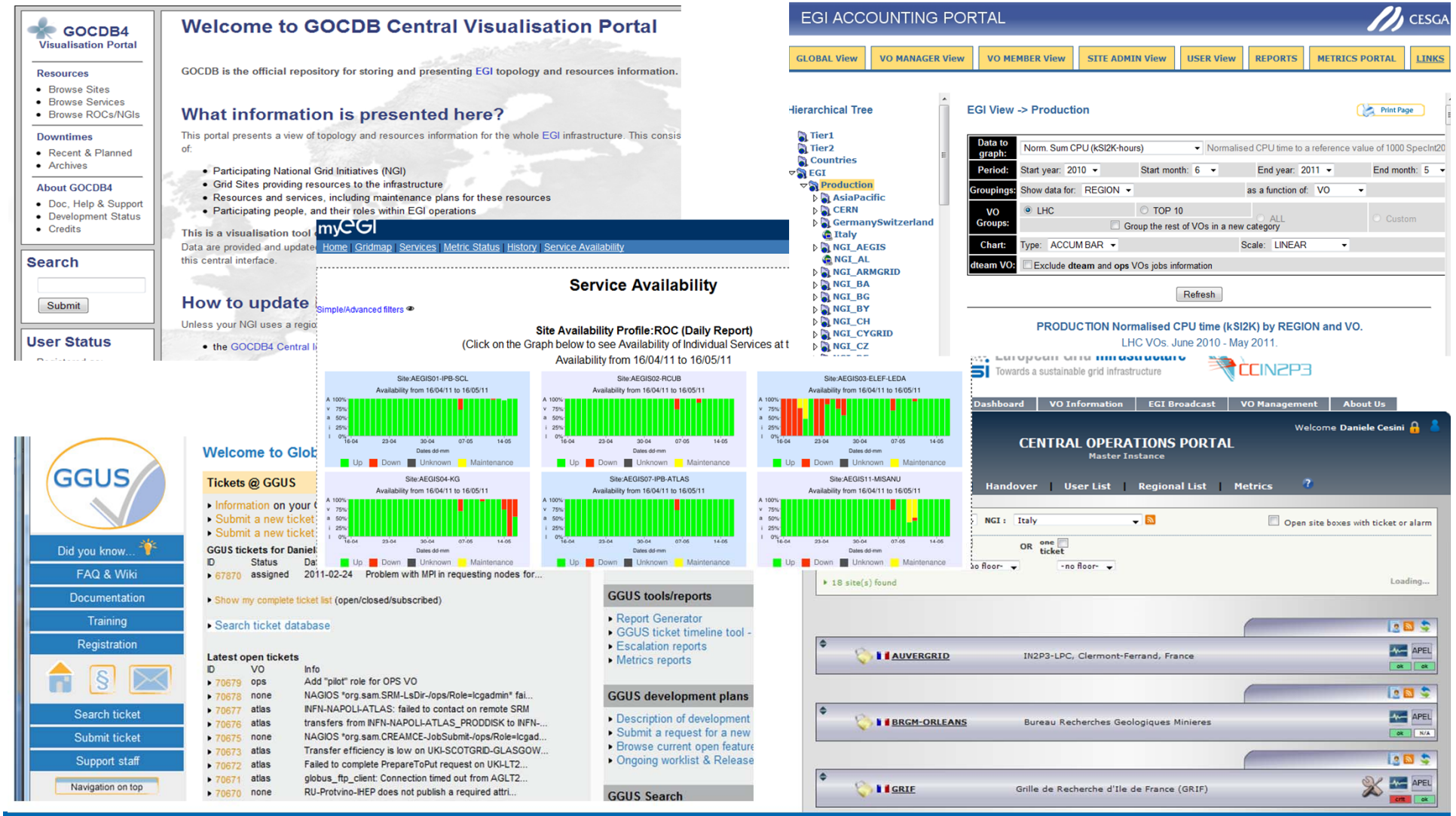
- **1 EIRO Operations Centre**: CERN
- **4 Non-European Operations Centres**
  - *Asia Pacific Federation*: Australia, China, India, Japan, Malaysia, Philippines, South Korea, Taiwan and Thailand
  - *Canada Federation*: Canada and China
  - *GISELA Consortium* (IGALC Federation): Argentina, Brazil, and Venezuela
  - *Latin America*: Brazil, Chile, Colombia and Mexico

Operational Tools

- Develops and maintains operational tools
  - Regionalised for NGI deployment

Tool	Link	Partner
Operations Portal	https://operations-portal.egi.eu/ The old CIC portal	CNRS
Service Availability Monitoring	Nagios and MyEGI portals fully regionalised at NGIs https://wiki.egi.eu/wiki/SAM_Instances	CERN/SRCE
Grid Configuration database (GOCDB)	https://goc.egi.eu/	SFTC
EGI Helpdesk (GGUS)	http://helpdesk.egi.eu/	KIT
Accounting Portal	http://accounting.egi.eu/	FCTSG
Metrics Portal	http://metrics.egi.eu/	FCTSG

Ops Tools Screenshots



Operational Tool Developments

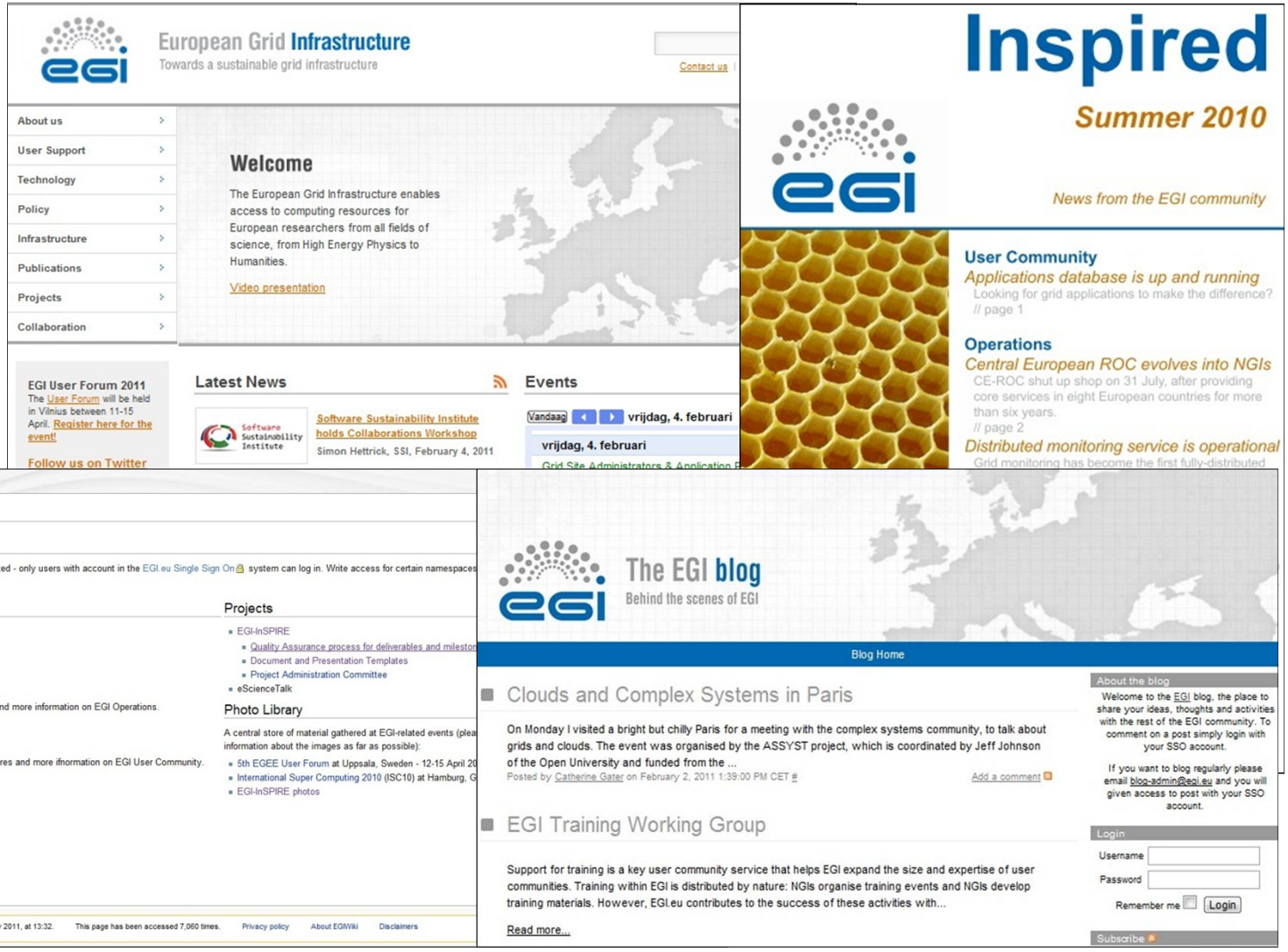
- Message Broker Network Configuration to support tools and operations
- Development of the Central Accounting Repository (clients provided by the EMI project)
- Development of accounting systems needed for “new” resource types
  - Billing
  - Accounting of application usage
  - Accounting of data usage
  - Accounting of capacity and cloud computing usage

Dissemination

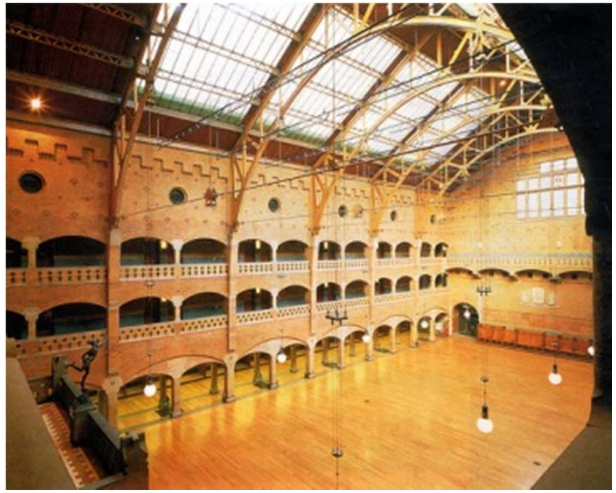
- Disseminate EGI's activity within the project and worldwide through its regionally dispersed dissemination contacts.
- Developed the EGI branding and content for the project and event websites.
- Produced 12 monthly Director's Letters, 4 issues of the EGI *Inspired* newsletter, published success stories in *International Science Grid This Week* and *Public Service Review*.
- Produced a range of brochures and posters.
- EGI website received over 30,000 visitors in its first year, around 365,000 page views.
- Attended a range of international events, including ISC2010 in Germany, ICT 2010 in Brussels, eChallenges in Warsaw, SciTech in Brussels, ISGC in Taipei, and SC10 in the US.
- Hosted booths at the EGI Technical and User Forums, and ran outreach campaigns that included printed materials, press releases, social media feeds and blogs.
- NGIs in the International task contributed events, websites, materials, publications, papers, translations, press releases and outreach to policy makers.

Communications

- Website
- Wiki
- Blogs
- Newsletters
- Letters
- Social Media



EGI Technical Forum 2010



- Held in Amsterdam at the Beurs van Berlage,14 - 17 September 2010 in partnership with the BiG Grid project, and attracted 570 delegates.
- First major event for the EGI community and brought together European distributed computing projects and their collaborators in academia and businesses.
- Aimed to establish collaborations between the new and the current European DCI projects to meet the needs and requirements of the research community.
- Included 290 contributions in the form of presentations, demos, posters and workshops.
- 22% of delegates used the iPhone application, 28% used Twitter, 10% Flickr, 25% YouTube, 8% the GLOBAL webcast of the plenaries and 28% read the GridCast blog.
- Two press releases were issued leading to 27 press cuttings, including *HPCwire*, *iSGTW*, *ZDNet*, *Yahoo News*, *ITnews* in Australia, *Science Business* and *Environment & Energy Management*.
- GridCast team from e-ScienceTalk produced 26 posts on the blog and 6 videos on YouTube.

EGI User Forum 2011

- Organised by EGI.eu, Vilnius University and LITNET in Vilnius, Lithuania, 11-14 April 2011, with the support of the European Middleware Initiative (EMI) and local secretariat BAIP.
- Held at the Radisson Blu Lietuva in Vilnius, Lithuania, showcasing the diversity of the EGI user community, attended by 427 delegates.
- Programme included networking and opportunities to ‘meet the EMI experts’.
- Featured 196 contributions, 173 speakers and 34 session conveners.
- Over 250 images on Flickr, 30 posts on the GridCast blog, including 14 videos and slide shows. Over 2,600 unique visitors visited the main event website, representing 20,000 page views.
- Press articles appeared in *HPCWire*, the SSI blog, the GÉANT newsletter and *iSGTW*.
- Book of Abstracts published for distribution at the event.

EGI means Innovation

- Deploy Technology Innovation
  - Distributed Computing continues to evolve
    - To include: Grids, Desktops, Virtualisation, Clouds, …
- Enable Software Innovation
  - Provide reliable persistent technology platform
    - Tools built on gLite/UNICORE/ARC/Globus
- Support Research Innovation
  - Infrastructure for data driven research
    - Support for international research (e.g. ESFRI)

Technology Innovation

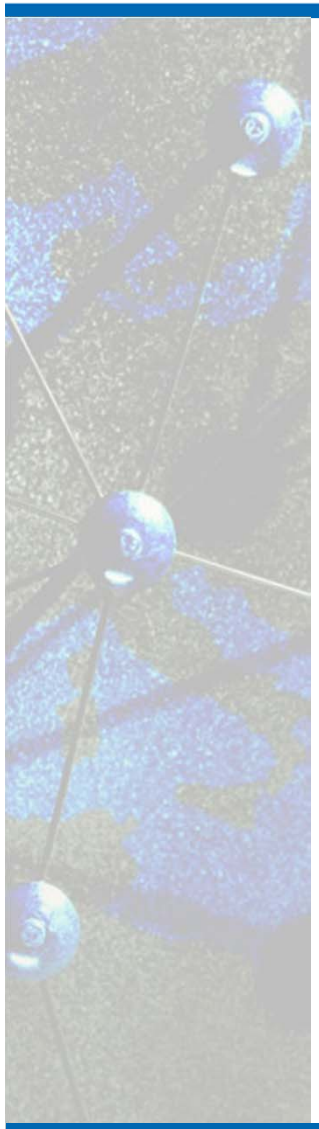
- Will come from outside EGI
  - Moving research technologies into production
- Partnership with technology projects
  - EMI (European Middleware Initiative)
  - IGE (Initiative for Globus in Europe)
  - EDGI (European Desktop Grid Initiative)
  - StratusLab
  - VenusC
  - SAGA

Software Innovation

- Will also come from outside EGI
  - EGI is a neutral platform for applications
- EGI cannot support all services for all users
  - Every community needs something different
- Foster innovation within different ‘sectors’
  - e.g. 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
  - Enable their community specific support activity:
    - Support, training, consultancy, requirements etc.



Grid Use Cases

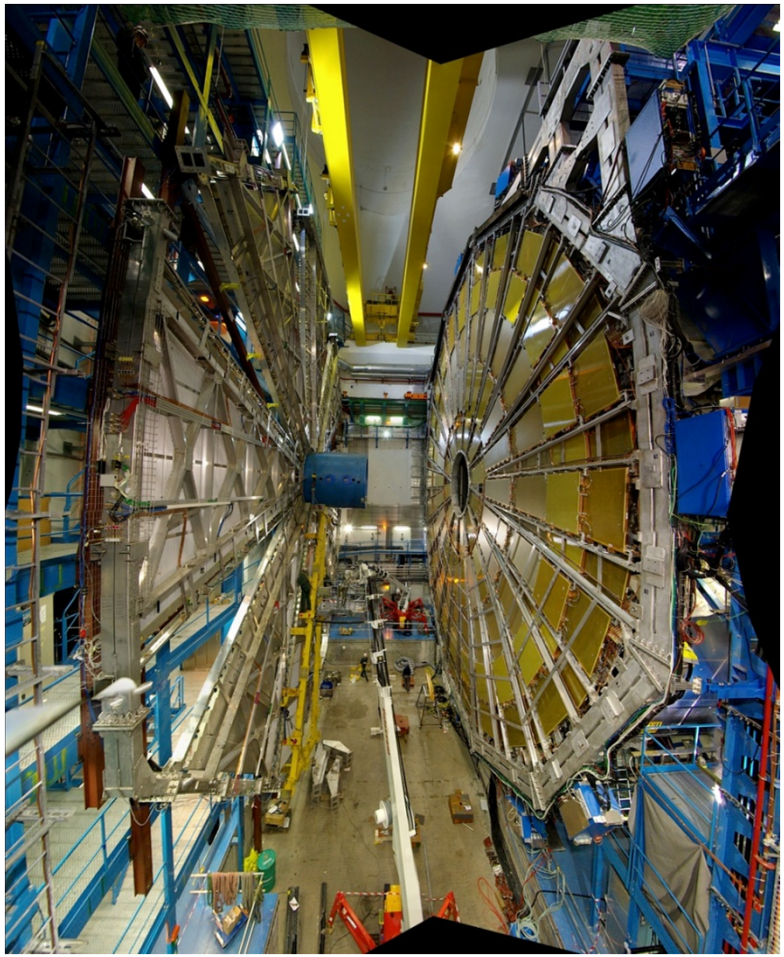
Grid Use Cases

These case studies show some of the advantages of using the grid:

- allows world-wide multi-disciplinary collaboration;
- integrate distributed resources into a single whole;
- customised grid services to meet the unique demands of researchers;
- reliable service for computation, data transfer and storage of large data sets;
- reduced analysis time and analysis on-demand;
- scientifically useful results are generated more quickly;
- long term support;
- sharing sensitive data securely among a trusted community;
- allows member institutions to contribute computing power to the community;
- generate data-intensive stimulations in a shorter amount of time;
- reduce technical workload (by following grid standards), so scientists can concentrate more effort on the science

Use Case: Large Hadron Collider

- World's largest particle accelerator
- Supports 8,000 researchers
- 1 billion CPU hours in the last 12 months
- 15Pb of data created annually



Use Case: Large Hadron Collider

Some advantages of using the grid:

- allows worldwide mass collaboration with thousands of physicists;
- customised grid services to meet the unique demands of the experiments;
- large data storage facility;
- physicists can access the data using their own computer locally.

Use Case: GoNL

- Mapping the genome of the Netherlands
- Plan to sequence the genomes of 750 Dutch people
- Currently 30Tb of data
- Will generate 20 times that amount

Use Case: GoNL

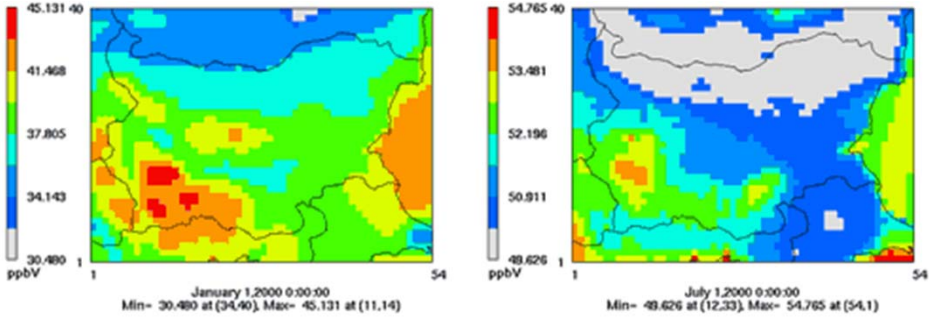
Some advantages of using the grid:

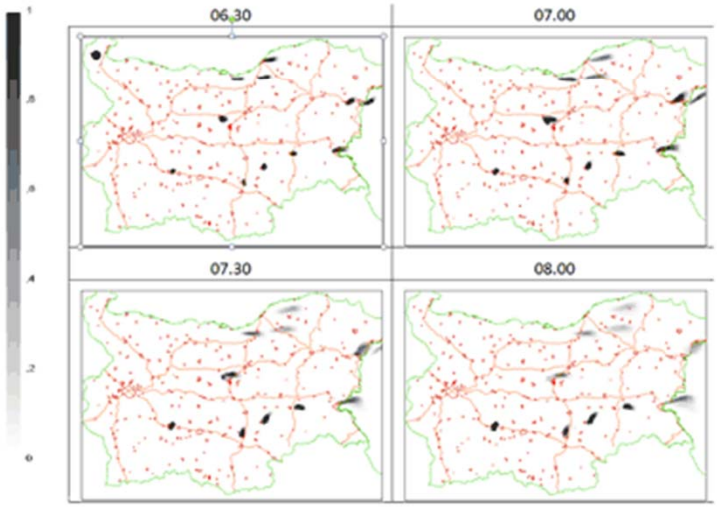
- analysis time reduced by 80%;
- on-demand analysis – can be carried out as and when researchers need it;
- provide development and support to help researchers get the most out of using the grid.

Use Case: Environmental Modelling

- Bulgarian researchers have ported three applications to the grid

1. Study the impact of climate change on air quality
2. Model atmospheric composition
3. Investigate emergency responses to the release of harmful substances into the atmosphere





Use Case: Environmental Modelling

Some advantages of using the grid:

- improved response times and decreased failure rate, so scientifically useful results are generated more quickly;
- reliable service for computation, data transfer and storage of large sets of data;
- using existing software with standard protocol means a quicker start-up time and compatibility between resource providers.

Use Case: ASTRA

- Ancient instruments Sound/Timbre Reconstruction Application
- Has recreated 4 instruments so far
- Held concerts using these instruments

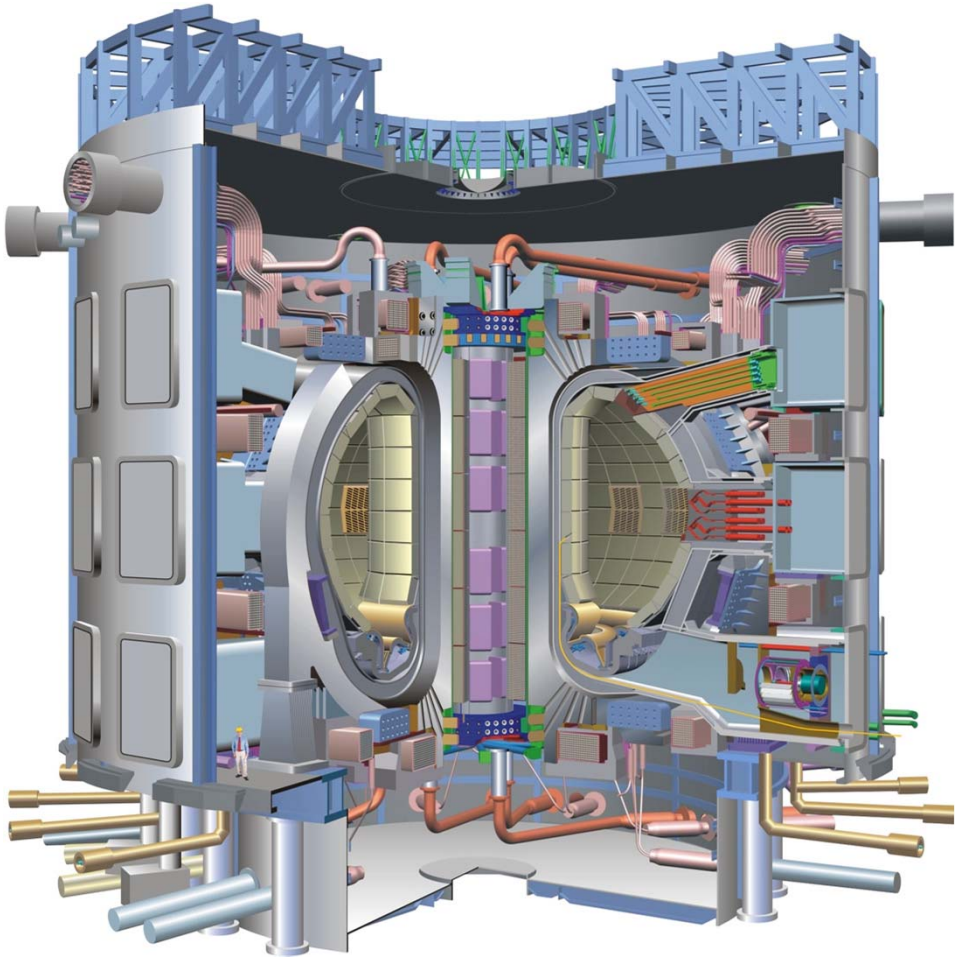
Use Case: ASTRA

Some advantages of using the grid:

- can meet high demand for network and computing requirements;
- high reliability;
- allow multi-disciplinary collaboration between researchers, musicians and historians;
- longevity: ASTRA running since 2006.

Use Case: ITER

- Investigating viability of fusion as a power source
- Modelling and simulating the reactor
- Used 1 million CPU hours in the last 12 months



Use Case: ITER

Some advantages of using the grid:

- perform the intensive computations needed to test the feasibility of fusion power before building the reactor;
- open to future development: dedicated project ‘EUPHORIA’ was set up to further push the limits of existing state-of-the-art computing resources.

Use Case: DECIDE

- Diagnostic Enhancement of Confidence by an International Distributed Environment
- Diagnostic tools for the medical community
- Example: Their Statistical Parametric Mapping application can help doctors to diagnose Alzheimer's disease in its early stages and track the progress of the symptoms over time

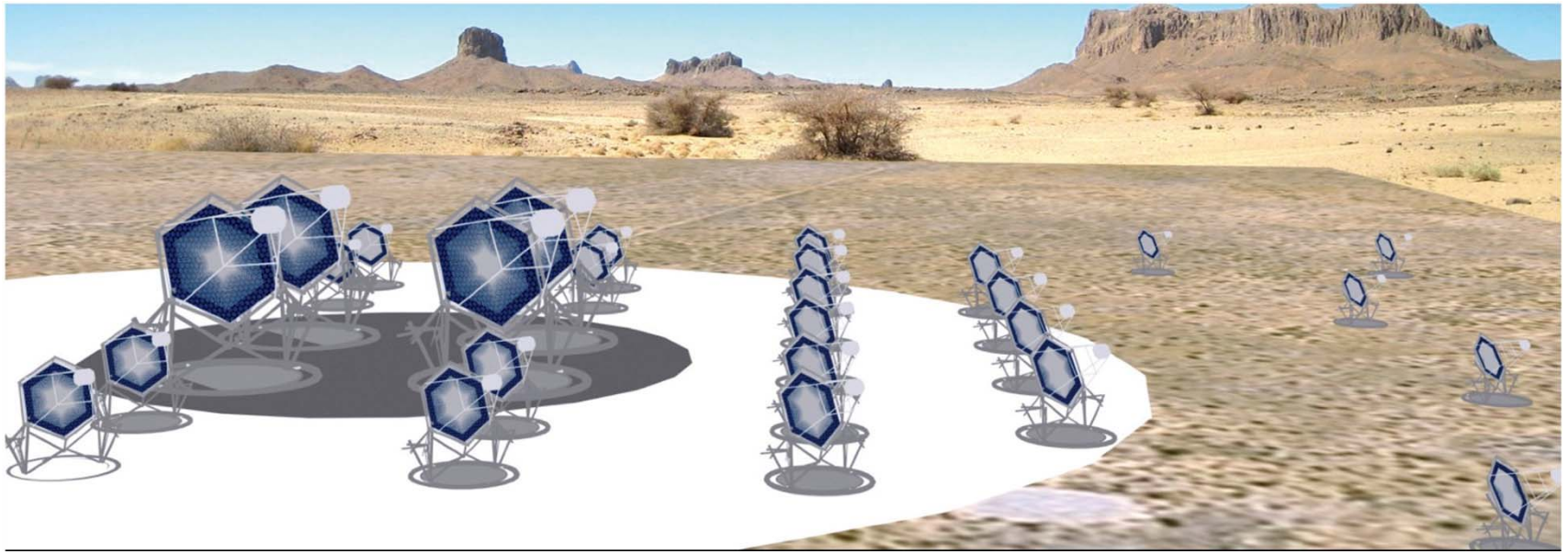
Use Case: DECIDE

Some advantages of using the grid:

- a single European-wide master database of images stored on the grid for doctors to use;
- can set up diagnostic tools with a dedicated grid infrastructure;
- customisable: dedicated software to track progression of the disease over time;
- sharing medical data securely.

Use Case: CTA

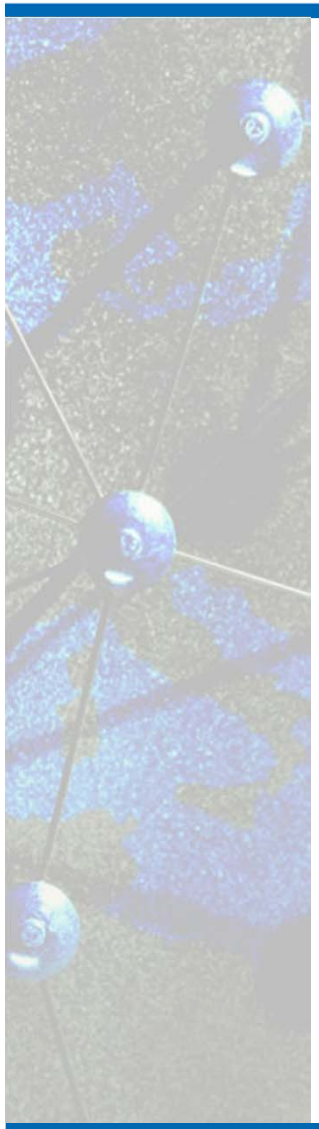
- The Cherenkov Telescope Array
- Future ground-based high energy gamma-ray instrument
- 132 institutes in 25 countries
- Using applications and grid technology provided by the European grid



Use Case: CTA

Some advantages of using the grid:

- allows member institutions to contribute computing power to the CTA community;
- generate data-intensive stimulations in a shorter amount of time;
- reduce technical workload (by following grid standards), so scientists can concentrate more effort on the science.



Future Plans

Learning 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
- The impact of commodity virtualisation…
  - For transactional models →
    - The ‘Cloud’: A model based on compute not data
  - For large distributed data-oriented models →
    - The emergence of true ‘function shipping’?

Guiding Principles

- Be a neutral resource provider
  - Any application, any domain, any technology
  - A platform for domain specific innovation & use
  - Integration of any compliant resource
- End-user needs and technologies change
  - Allow VOs to deploy their own services
    - VOs will then need to manage **their** infrastructure
  - Give VOs the power to meet their own needs

Standards Needed Here!

- Data Layer
  - Secure reliable data movement
  - Access to data resources
- Virtualisation Layer
  - Span trust domains within agreed policies
  - Monitoring as important as lifecycle control
- Service Layer
  - The services that go into the virtual machine
  - Avoid domain specific silos & promote reuse

Consensus
Openness
Balance
Transparency

What will EGI do?

- Continue with a secure reliable infrastructure
  - Integrated: gLite & ARC
  - Underway: Globus & UNICORE
- Support its user communities
  - Maintain user services & tools
  - Engage with structured (virtual) user communities
    - Encourage structuring in unstructured user communities
    - Defined representatives within EGI bodies
  - Engage with the ESFRI projects

Summary

- EGI.eu established in Amsterdam
  - Supported through EGI-InSPIRE project
- EGI has transitioned from a federation of regional to national resource providers
- EGI will with technology providers to have a standards based open architecture
- EGI will evolve to support the needs of its current and new user communities

Contacts

- EGI.eu Director
  - director@egi.eu
- EGI.eu Operations Team
  - operations@egi.eu
- EGI.eu User Community Support Team
  - ucst@egi.eu
- EGI.eu Policy Team
  - policy@egi.eu
- EGI.eu Dissemination Team
  - press@egi.eu
- EGI.eu Secretariat
  - contact@egi.eu