



Analysis of ESFRI projects' e- infrastructure requirements by the European e-Infrastructure Forum

Bob Jones, CERN

bob.jones@cern.ch





EEF - European E-Infrastructure Forum

- **Forum for the discussion of principles and practices to create synergies for distributed Infrastructures**
- **Goal:** seamless interoperation of leading e-Infrastructures serving the European Research Area
- **Focus:** needs of the user communities that require services which can only be achieved by collaborating Infrastructures
- **Initial membership:**
 - EGEE & EGI
 - DEISA & PRACE
 - Terena & GEANT
 - The forum recognises the importance of data access and management and is seeking to add a member specialising in service provision in this area for multiple research communities
- **Offers** a way of interacting as a whole with user communities of a multi-national nature that are interested in making use of the Infrastructures

<http://www.einfrastructure-forum.eu/>





Objectives

- Provide input on the vision and model for Europe's e Infrastructures
- Offer a clear added value to key user communities by providing access to Europe's e Infrastructures in a coordinated manner and help them devise computing models that make use of the European e Infrastructures
 - Share information within the group on status and direction of e Infrastructures
 - Gather input and form common opinion on subjects that will impact the long term goals, function and effectiveness of Europe's e Infrastructures
 - Present a common view on European e Infrastructures to peer groups elsewhere in the world
 - Interact with emerging structures to transmit the experience gained from existing production Infrastructures and formulate common user requirements





EEF - European E-Infrastructure Forum

Requirements Questionnaire

- **Small set of questions give overview of e-Infrastructure requirements**

<https://documents.egi.eu/public/RetrieveFile?docid=9&version=1&filename=eef-questionnaire.docx>

- Need technical contact(s) to discuss in further detail

- **Questions**

- ICT vision - how Information and Communication Technologies contributes to the overall research facility vision and goals
- Major ICT requirements (Storage, Compute, Networking, Security & Access)
- ICT architecture
- Current Status & Timeline
- Challenges and suggestions

- Questionnaires completed by a number of ESFRI projects
- **First step in what EEF believes will be an iterative process**





EEF - European E-Infrastructure Forum

Workshops at which EEF has gathered further information about ESFRI project requirements:

- EGEE09 (Barcelona, Sept'09)
- 7th concertation event (Brussels, Oct'09)
- NEERI09 (Helsinki, Oct'09)
- ERF (Lund, Oct'09)
- BMS (Brussels, Dec'09)
- SSH (Brussels, Jan'10)
- ENV (Brussels, Mar'10)
- ECRI2010 (Barcelona, Mar'10)
- EBI (Cambridge, Apr'10)
- EGEE user forum (Uppsala, Apr'10)

Total of 28 ESFRI (+ eNMR and NeuGrid) projects consulted

What follows is a first pass analysis of these requirements

Report published: <http://www.einfrastructure-forum.eu/documents/EEF-report>





Common SSH project requirements:

- Data archiving and curation
- Flexible repository system with support for Persistent Identifiers (PIDs)
- Sensitivity of the data calls for a fine grain Authorization and Authentication system
- Access to grid and cloud computing facilities
- Ease of access to all e-infrastructure resources via Single Sign On
- Education and training for e-infrastructure usage





Common BMS project requirements:

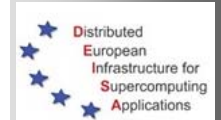
- Ability to manage massive quantities of data that is generated across Europe and continuously growing
- Access to a wide variety of databases which are regularly updated
- Specific sub-communities deal with patient data where secure data management is essential
- Ease of use via Single Sign On facility
- Large-scale usage of grid and cloud computing
- Collaboration with partners in USA and Japan is essential
- Education and training for e-infrastructure usage





Common ENV project requirements:

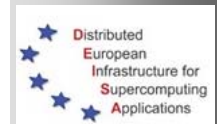
- Mixture of fixed and mobile facilities
- Ability to gather data from large-scale & distributed sensor networks in a timely manner (some real-time constraints)
- Combining, processing and storage of large, complex datasets





Common Energy project requirements:

- Primary source of e-infrastructure requirements has been the fusion community
 - example where they have already demonstrated (via EUFORIA project) using multiple infrastructures in the same work-flow (GEANT, EGEE, DEISA)
- Emphasis on modeling, simulation and processing facilities for single-site installations
- Access to HPC and grid systems with the ability to run complex work-flows across the range of e-infrastructures
- Education and training – notably application code porting to grids and massively multi-core machines





Common Material and analytical facilities & Physical and engineering sciences project requirements:

- Emphasis on modeling, simulation and processing facilities for single-site installations
- Data curation, archive and distribution
- High-speed, reliable networks for continuous data transfer
- Open access to data and results to large, widely distributed user communities across and beyond Europe
- Single sign-on service





First pass analysis of ESFRI requirements (all sectors) identified the following common areas

- Single sign-on – consistent access to resources
- Virtual organisations (collaboration)
- Persistent storage – long term preservation of data and its access
- Data Management services
- Standards – web services
- Workflows – support of access to HPC/grid/network resources (compute & data) across Europe for the
- Training & consultancy
- Global scope – beyond Europe





EEF - European E-Infrastructure Forum

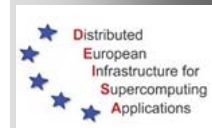
EEF thinks it can address these common requirements by harmonising existing services so users get a consistent access to all e-infrastructure resources:

- Single sign-on – ensure same identify can be used across network/HPC/grid by harmonising policies for Authentication, Authorization and eventually Accounting and Auditing
 - needs technical developments and a review of policies by all parties
- Virtual organisations – harmonise support across HPC & grids
 - review policies and potentially some middleware developments needed
 - Security incident handling (e.g. cooperating security incident response group)
 - integrated security groups



Harmonising existing services (2/3):

- Persistent storage
 - work with middleware & data mgmt providers so PIDs are supported
- User support (e.g. interoperating problem handling procedures)
 - Training & consultancy (contribute to customised training events/material organised by ESFRI projects)
 - EGI relies on NGIs for material and trainers
- Standards – offer web service interfaces for all relevant e-infrastructure services
 - will require some middleware developments





Harmonising existing services (3/3):

- Workflows – support of access to HPC/grid/network resources (compute & data) by a variety of workflow engines
- Global scope – beyond Europe: leverage existing EEF connections and contacts with sister e-infrastructures around the world
- cloud / volunteer computing integration
 - rely on developments by 3rd party projects (StratusLab, EDGI etc.)





Data Management

- All ESFRI projects identified data management as a requirement
 - Large variation in the interpretation of what data management implies
 - data transfer
 - data sets (simple files) and databases
 - ontologies for categorization and structuring of data
 - metadata
 - range of data related standards referenced
- EEF has an “empty seat” reserved for European data services provider(s)
 - GRDI2020 identified 12+ existing data related EC projects
- It is assumed that the European data services provider(s) would address advanced data management requirements and existing EEF members would ensure such services can exist and be supported in the European context





Next Steps

- EEF welcomes feedback from all stakeholders on the report
- This is an iterative process – expect to refine the requirements analysis and planning taking into account feedback and additional input from more ESFRI projects and user communities
- The actions required cannot all be accommodated in the programme of work and resources of the e-infrastructures (DEISA/PRACE, EGI-InSPIRE, GEANT3)
- A number of recently started projects may be able to contribute (EMI, StratusLab, EDGI etc.)
- Identify joint projects with ESFRI projects (already started within BMS and SSH) and extends to all ESFRI sectors
 - Must avoid scenario where each ESFRI project adopts a model which is unique

