GridTalk GridBriefings
Grid computing in five minutes

New projects, new requirements: Future needs of the ESFRI projects

The European Strategy Forum on Research Infrastructures (ESFRI) is a European Commission initiative which supports the long-term development of Europe's scientific research infrastructures. Projects in the ESFRI roadmap span social and biomedical sciences, earth and physical sciences, energy, infrastructures and analytical facilities. With such wide scope and vision, the ESFRI projects will place large demands on the storage, processing and networking services of Europe's e-Infrastructures. In April 2010 the European e-Infrastructure Forum released a report that examined the requirements of the ESFRI-projects and outlined the services and resources that the e-Infrastructure community can offer. This briefing summarises that report and the issues it raises.

What is ESFRI?

ESFRI was formed in 2002 at the request of the European Council. ESFRI's delegates are nominated by the Research Ministers of the Member and Associate Countries, and also include a representative of the Commission. Its role is to guide policy making on Research Infrastructures (RIs) in Europe.

Tasks carried out by ESFRI include identifying gaps in Europe's research capacity, giving recommendations on the management of RIs and communicating the importance of RIs to the wider public. ESFRI does not itself fund RIs but hopes that its work will help meet the needs of the scientific community, and keep Europe at the forefront of research.

The ESFRI Roadmap, last updated in 2008, identifies vital European RIs of different size and scope which correspond to long term needs of the European research communities. In October 2009 ESFRI invited the e-Infrastructure Reflection


GridTalk

Material and analytical facilities & physical and engineering sciences

The material and analytical facilities ESFRI projects have been grouped with those of physical and engineering sciences due to the closely related nature of the RIs concerned. Areas covered include high magnetic fields and sources of photons and neutrons.



European XFEL

Example project: The European **XFEL** (the European X-ray Free-Electron Laser) being built in Hamburg and Schleswig-Holstein, Germany, will be a world leading facility for the production of intense, short pulses of X-rays for scientific research in a wide range of disciplines.



Krzysztof Wrona, European XFEL – *"The European XFEL will generate data volumes that will require large, secure data storage and archive systems supporting fine-grained access protection mechanisms. The processing model proposed requires high data throughput computing facilities that are closely coupled to data storage. Reliable WAN connections are required to allow users access to data and processing facilities via single sign-on and user-friendly services such as web portals, workflow management tools and software repositories."*

Requirements for this sector:

- A need for high performance modelling, simulation and processing facilities.
- Projects require reliable high-speed network connections for data acquisition as well as for distributing data to a large number of researchers around the world.
- This sector also needs data archiving facilities and open access for the user communities, for whom a single sign-on service will be vital.

Group (e-IRG) to produce a report, or "Blue Paper", on e-Infrastructure services, to be used as input for the 2010 update of the ESFRI roadmap. This e-IRG Blue Paper will be formally delivered to ESFRI at the end of June 2010. It provides an assessment of Europe's e-Infrastructure service portfolio and trends, and sets out policy recommendations for several key areas.

Social sciences and humanities

ESFRI projects in this area include data archiving, social surveys, and language resources and technologies, as well as digital humanities and cultural information.



Example project: **The European Social Survey** (ESS) monitors the changing attitudes, beliefs and behaviours of people in over 30 countries.



Roger Jowell, ESS: *"In preparation for its upgrade to a European Research Infrastructure Consortium (ERIC), the ESS has been developing its e-Infrastructure strategy for the long-term. The ESS Data website offers on-line analysis, visualisation of data and the opportunity to download all or part of the datasets. Issues of metadata, interoperability, preservation and enhanced dissemination tools, common to other e-Infrastructures, are also central themes in the ESS ERIC e-Infrastructure strategy."*

Requirements for this sector:

- Projects need a flexible repository system with long-term means of identifying data (called Persistent Identifiers) to provide the basis for data storage, archiving and management.
- The sensitive nature of the data requires a fine-grained authentication and authorisation system.
- Single sign-on functionality is important.
- Some projects plan to use grid or cloud compute facilities to process stored data.
- Education and training on e-Infrastructures and associated technologies are required.

What is the e-IRG?

The e-IRG is an inter-governmental policy body comprising national delegates from more than 30 European countries. It defines and recommends best practices for pan-European e-Infrastructure efforts.

What is the European e-Infrastructure Forum (EEF)?

The EEF is a forum for European e-Infrastructure providers. Members of the EEF cover:

- Grids: **EGI**, the European Grid Infrastructure.
- Supercomputing: **DEISA**, Distributed European Infrastructure for Supercomputing Applications, and **PRACE**, Partnership for Advanced Computing in Europe.
- Networking: **Terena**, the Trans-European Research and Education Networking Association and **GÉANT**, the pan-European data network for the research and education community.

Through a series of meetings and a questionnaire, the EEF gathered a set of e-Infrastructure requirements from 28 ESFRI projects. The resulting report aims to tailor EEF services to the ESFRI projects' needs.



Bob Jones, EEF *chairman – "This EEF report is just the first step. Now, through a combined effort of the existing e-Infrastructures, ESFRI and collaborating projects, we have to build a seamless ecosystem to serve Europe's research communities."*

Challenges and opportunities for e-Infrastructures

The EEF report found that many ESFRI projects will need to make use of existing e-Infrastructures, so improving their interoperability will benefit all user communities. The section below outlines common ESFRI requirements along with the implications and opportunities these represent for EEF members.

Technology

Integration with cloud and volunteer desk-top systems: Most of the ESFRI projects consulted are unconcerned about where resources come from, they simply ask for easy-to-use, yet powerful and secure, data management facilities. Each computing paradigm has advantages and drawbacks, so a custom-fit solution for each community is likely to work best. Standards bodies, such as the Open Grid Forum, aim to improve interoperability between clouds, grids and supercomputing.

Persistent storage: Long-term storage and accessibility was a requirement of several sectors, and Persistent Identifiers and metadata are key issues for user communities. Effective access to persistent data from the European e-Infrastructures implies:

- Centres offering persistent data will need to guarantee good quality of service and access.
- Middleware deployed by e-Infrastructures will have to

Biological and medical sciences

These projects cover a range of disciplines with a general focus on health and drug development. There is also some work in the area of marine biology.

Example project: **ELIXIR** (European Life-Science Infrastructure For Biological Information) will build a distributed but interlinked collection of biological data resources and literature for research in the life sciences as well as related research in medicine and the environment, the bio-industries and society.



Janet Thornton, ELIXIR - *"Europe is facing unprecedented challenges in healthcare, agriculture and environmental protection. Research in these areas is generating vast quantities of data about biological molecules that need to be curated, archived, interpreted and made available to researchers across the life sciences and medicine. ELIXIR is constructing an infrastructure across Europe to implement this to ensure that Europe remains at the forefront of science in this vital area. Clearly this will only be possible with the cooperation of the European e-Infrastructure providers and we are already working with members of the EEF to ensure that this happens."*



APO - Photodisc

Example project: **BBMRI** (Biobanking and Biomolecular Resources Research Infrastructure) is building a distributed pan-European infrastructure of bio-banks in an effort to identify new targets for therapy as well as for drug discovery and development.



Jan-Eric Litton, BBMRI - *" Many countries have recognised the importance of biobanking, and there are many initiatives establishing national biobanks. BBMRI's distributed "hub and spoke" structure provides great flexibility so that new members can be connected at any time and it can easily be adapted to emerging needs in biomedical research. The IT infrastructure employs federated database architecture and grid-computing technology, and will integrate the complex network of hubs, members and partners into a single virtual infrastructure."*

Requirements for this sector:

- Huge increases in the amounts of data available require well structured databases, which should be broadly accessible.
- The value of the data being collected far exceeds the costs of storing and accessing it, so developing a distributed infrastructure to store, curate and provide global access is key.
- For user-friendly data access, major investment in computer infrastructure and storage is envisaged, along with the development of appropriate standards and ontologies.
- Developments in imaging are likely to give rise to significant increases in data volumes, placing new demands on computing, networking and storage.
- The main demands on e-Infrastructures are in terms of storage, grids, networks and general computing.

be modified to support access to persistent data using Persistent Identifiers at these sites.

- Being able to record and trace the origin of data as well as its movement between databases.

Single sign-on: A unified single sign-on service would ensure an individual's identity can be used across all EEF services. Existing EEF Authentication and Authorisation Infrastructures (AAI) are similar but not identical, and are separately managed by different e-Infrastructures. The ESFRI projects require existing AAI systems to interoperate so that a user's identity can be established once and accepted by all.

Workflows: The need for workflows was identified by all ESFRI research sectors. There are many workflow tools or frameworks employed by user communities and this diversity is certain to remain. EEF members will work to support a range of workflow tools and environments and allow seamless access to cross infrastructure workflows.

Users

User support: To use European e-Infrastructures effectively, users should receive quick responses to questions as well as high-quality documentation. All large-scale European e-Infrastructures offer specific user support facilities. EEF members will need to ensure these existing support structures cooperate, so users can issue requests to appropriate support groups across different technologies and separate geographical regions.



Hermann Lederer, DEISA – *"Identifying the common IT needs of ESFRI projects enables exploitation of synergies and services already put in place by leading European e-Infrastructures in the areas of high speed networks, capacity computing and capability computing. Through DEISA, we can offer services that will be used by the leading European HPC systems."*



Thomas Eickermann, PRACE – *"Many Research Infrastructures, rely or will rely to some extent on resources and services provided by PRACE and other e-Infrastructures. The EEF survey is an important first step to identify commonalities as well as differences among the ESFRI projects' requirements. It will help the EEF to set priorities for its work towards a seamless interoperation of the e-Infrastructures."*

Training and consultancy: All ESFRI projects have expressed a need for training, education or external expertise to be able to use e-Infrastructures. EEF members offer varying levels of support in this area. Experience shows that training programmes have the most impact when tailored to the specific needs of the target user community. EEF members are willing to contribute trainers and material to training programmes and events that could be organised by the ESFRI projects, either individually or on a sector basis.

Virtual Organisations: HPC and grid currently support and provide for Virtual Organisations in vastly different ways. A goal of EEF members will be to offer consistent support for Virtual Organisations by harmonising current features.

Web-service interfaces: A common theme among ESFRI projects is the desire for services that can be used directly on the World Wide Web as a standardised manner of packaging e-Infrastructure services. Where appropriate, e-Infrastructures should offer web interfaces for relevant services and allow user communities to build on these to produce their own customised gateways.

Policy

Global scope: All ESFRI sectors need to collaborate with parties outside Europe. EEF members can use their existing international contacts for the benefit of the ESFRI projects. Service availability for ESFRI projects should not be a problem but in some locations, access to services would need to be confirmed and addressed.



Steven Newhouse, EGI – *"The EEF report has identified some of the critical functional building blocks needed by the ESFRI projects to accelerate their adoption of European e-Infrastructure. This will help the European e-Infrastructure providers prioritise their joint activities, aiming to provide an integrated e-Infrastructure for the European Research Area."*



Karel Vietsch, TERENA - *It is essential for research networking organisations to understand the needs of their users. At the European level, the GÉANT project has collected networking requirements from large international projects, as reflected in the EEF report. Europe's global leadership in middleware, supported by TERENA, will help address single sign-on needs.*

Environmental sciences



These projects focus on climate, ecological, atmosphere, earth and ocean systems. They support the sustainable management of the environment by monitoring and measuring major environmental systems.

Example project: **IAGOS** (In-service Aircraft for a Global Observing System) measures atmospheric composition by installing scientific instruments on commercial aircraft. Its database can be used for weather prediction, air quality forecasting and in climate and carbon cycle models.



Andreas Volz-Thomas, IAGOS-ERI – *"To provide in-situ information on atmospheric composition, both in real-time and for long-term scientific usage, appropriate e-Infrastructure is being developed in IAGOS in addition to utilising existing structures of the meteorological services. Additional structures that put the information into context for users beyond the scientific community, e.g. Global Earth Observation System of Systems (GEOSS), are most beneficial if developed in close cooperation with data providers and users."*

Requirements for this sector:

- Projects require measurement and monitoring facilities, access to analytical facilities such as synchrotrons, as well as large-scale access to global and distributed facilities on a pan-European basis.
- Significant investment is proposed in both fixed and mobile server systems, which will collect data from land, sea and air measurements.
- ICT challenges include data capture, particularly from sensor networks, and combining, processing and storing large and complex data sets.
- Significant real-time requirements are foreseen in terms of collecting and processing data.

For more information:

EEF: www.einfrastructure-forum.eu
e-IRG: www.e-irg.eu
ESFRI website: http://tinyurl.com/RIs-ESFRI
ESFRI 2008 Roadmap:http://tinyurl.com/ESFRI-roadmap08
iSGTW: www.isgtw.org
GridTalk : www.gridtalk-project.eu

GridTalk is co-funded by the EC under FP7