

Science, Strategy and Sustainable Solutions, a Collaboration on the Directions of E-Infrastructure for Science

A proposal for a cross-disciplinary user forum to collaborate on requirements and policies concerning e-infrastructure for science.

Contents

| | |
|--|---|
| Introduction | 2 |
| The Problem | 2 |
| E-Infrastructure Activities | 3 |
| Membership..... | 4 |
| Objectives | 5 |
| Deliverables..... | 6 |
| Recent Examples of User Coordination | 7 |
| FIM4R..... | 7 |
| Helix Nebula | 7 |
| Joint development with industry | 7 |
| Relationship with other Activities | 8 |
| Research Data Alliance | 8 |
| eInfrastructure Reflection group (eIRG)..... | 8 |
| Process..... | 9 |
| Next Steps..... | 9 |

This document has been prepared by David Foster & Bob Jones, IT dept. CERN on behalf of the EIROforum IT working group.

Introduction

As the importance of European e-infrastructures grows and matures it becomes increasingly important that user communities are able to voice requirements and help drive the direction of their evolution. However, the diversity of communities and the relative maturity in their international collaborations makes this difficult to implement in an efficient and cost effective manner.

Pan European e-infrastructure providers need to understand the requirements of a wide variety of possible communities and have contact with them. In general this is not the end Scientist or Engineer but the institute or project that supports them.

What is proposed here is a pan-European forum for organisations and projects that operate at an international level in order to present to the policy makers and the infrastructure providers where there are common needs and opinions and where there is divergence. This will enable both policy makers and e-infrastructures providers to have a view across many research domains and be able to take strategic decisions that will reflect the commonalities, and differences, that exist.

This forum should be independent of any supplier and engage across research domains so that shared information and strong representation can be established.

The Problem

E-infrastructure investments on-behalf of the publicly funded research community represent an important and growing budget item. As identified in the Geant Expert Group report, the user communities will increasingly be called upon to pay for the services they receive if e-infrastructures on which users can depend are to continue to survive. E-infrastructure costs will be an integral part of the cost of doing science and consequently e-infrastructure investments must make a substantial and sustainable impact in order to be justified. To evaluate the impact, it is essential that the market of end-users is well understood by funding agencies and e-infrastructure services providers. An unpopular way of describing this is that there must be a business case for an e-infrastructure, but indeed it is precisely the considerations of who will pay what for the services that needs to be addressed. This issue was highlighted by the e-Infrastructure Reflection Group (eIRG) in its 2012 roadmap report where it stated there is a “Lack of business models based on secure and sustainable funding streams for the use and innovation of e-Infrastructures”.

This is a difficult topic as more and more IT services are commoditized and available on the open market and the research communities will adopt the most cost effective options for them.

E-Infrastructure services provided to the research community then must be innovative and either address needs that are clear or create opportunities for evolving the “business” of science. In either case they must be relevant enough to attract the investment of the user communities that they target.

This body intends to drive this process by providing information that helps e-infrastructure providers to create solutions that are of value to the user communities and develop the business case.

E-Infrastructure Activities

There are currently a number of infrastructure activities that engage with communities in a variety of ways. Recently the GEANT project (GN3+) has reviewed its governance model and has concluded that it needs better representation from user communities. This awareness has been a growing trend across all providers.

| | |
|---------------|---|
| e-IRG | Engages with country representatives concerned with e-infrastructure policy. |
| EGI | Engages with National Grid Initiatives (NGI's) |
| GEANT | Engages with the National Research Networks (NREN's) |
| TERENA | Engages with National Research Networks, large users and commercial providers |
| PRACE | Engages with the Supercomputer users. |

The eIRG 2012 roadmap highlighted the need for *“a single organisation with a central role for user communities with a particular emphasis on involving large, advanced and well-organised user communities at a European level and beyond.”* and that *“On the strategic level user communities will have to organise themselves to drive the long-term strategy.”*

The purpose of this proposed user community body is to provide timely, relevant and coordinated information to all the above, and any future activities, in order to synthesize the trends, discontinuities and emerging needs of the European Scientific and Technical community.

Membership

It is the intention for this activity to be lightweight, open and have no rigid constraints. In particular it is important to understand the issues facing emerging and future activities in the science and technology domain such that policy makers can react accordingly and e-infrastructure providers can have a coordinated response.

To have a universal representation would lead to an unmanageably large group being formed with practical consequences on the activities that can then be performed.

The first approach will be to solicit individual membership from a limited but representative set of activities:

- The ESFRI cluster projects (BioMedBridges, CRISP, DASISH, ENVRI)
- The EIROforum members
- The new Flagship Projects for Horizon 2020 (Human Brain Project, Graphene)
- The ERF (European Association of National Research Facilities) that will provide representation of national research infrastructures across Europe
- The League of European Research Universities⁵
- The Association of European Research Libraries⁶

The intention is that the members should be able to represent the needs of the end-users supported by these organisations and projects. In addition to this standing body, a number of Open User Forums could be organized as widely publicized events where the work can be presented and additional input solicited from a wider community. Such events could be co-located with existing e-infrastructure or research community events to maximize engagement.

It is essential that the members are able to represent a strategic view and that a broad range of disciplines is represented.

Each member will be charged with identifying a number of individual researchers that are active in their discipline and not engaged with major research organisations. These individual researchers will represent the views of what is referred to as the long tail of science, namely important research being performed by small groups or establishments. These individuals will be invited to participate in the Open User Forums.

Members are expected to contribute in a written and oral fashion to the activities.

⁵ <http://www.leru.org/index.php/public/home>

⁶ <http://www.libereurope.eu/>

Objectives

The intention of this body is to coordinate the discussions broadly across organisations that have growing, or emerging, needs in the e-infrastructure space. The initial areas of common interest are expected to be:

- Networking
- Cloud
- Big Data

This is expected to change over time as new paradigms emerge and become important. Such strategic topics concerning e-infrastructures include, but are not limited to:

- The growth of demand and expectations from e-infrastructures in terms of infrastructure and services.
- The value of a proposed service or infrastructure to the scientific community.
- Interoperability and Sustainability
- Identifying inhibitors to use, including regulatory, procurement and legal as well as technical issues.
- Sharing of best practices and successful approaches.
- Collaboration and creation of common services.
- Creation of common understanding between service providers and the user communities.
- Several ESFRI projects have highlighted the value of “user aggregation” of needs with respect to working with industry. CRISP ran a workshop on how the research labs can work together with the IT industry more closely⁷.

eIRG in its blue paper from 2010⁸ includes a list of specific subjects that came out of the survey of the ESFRI projects performed by the European E-infrastructure Forum⁹:

- 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 and data) across Europe
- Training
- Global scope: beyond Europe

This body is not a technical body but will provide input to solution providers.

⁷ <http://www.isgtw.org/feature/next-generation-scientific-computing>

⁸ http://www.e-irg.eu/images/stories/eirg_bluepaper2010_final.pdf

⁹ <http://www.einfrastructure-forum.eu/>

Deliverables

This body will provide input to e-infrastructure providers and projects, the European Commission and national funding agencies as well as scientific communities themselves. In particular it may be expected to:

- Prioritise and publish issues facing the scientific communities in the areas of e-infrastructures.
- Maintain a database of contact information.
- Provide an estimation of the impact of the e-infrastructures on the research communities.
- Provide information on the potential for a service in terms of market size and likely adoption.
- Organise representative input from the scientific communities through workshops and polls.
- Participate in strategic discussions with e-infrastructure providers and projects.
- Participate and provide input on strategic directions from the scientific community for the eIRG, European Commission and national funding agencies.

Recent Examples of User Coordination

A number of recent examples, described below, show the value of user coordination with respect to eInfrastructure planning as proposed in this document.

FIM4R

The subject of secured access to e-infrastructure resources has been developed under the FIM4R group as a concrete example of what can be achieved:

A series of workshops started in summer 2011 to investigate Federated Identity Management for Research (FIM4R) collaborations. These workshops were started as an initiative by the EIROforum IT working group. Through these workshops, several research communities have converged on a common vision for FIM, enumerated a set of requirements and proposed a number of recommendations for ensuring a roadmap for the uptake is achieved¹⁰.

The authors of the paper come from different research communities that span all four ESFRI cluster projects (BioMedBridges, CRISP, DASISH, ENVRI).

The GEANT project has taken the contents of this paper and used it form part of its new project, GN3+, funded by the EC that started on 1st April 2013.

So here is a concrete example of how the research communities working together to clarify their requirements can provide strategic direction to the e-infrastructures. The FIM4R workshops have also spurred a series of pilot projects that jointly involve the research communities and e-infrastructures which are being used to explore the requirements and check the suitability of the services being offered by the e-infrastructures.

Helix Nebula

Similarly, support of access to HPC/grid/network resources (compute and data) has also seen some concrete results. ESA, CERN and EMBL have come together to specify their requirements for the use of commercial cloud computing services and consequently formed the Helix Nebula initiative¹¹ as a public private partnership with a growing number of suppliers. GEANT and EGI have joined the initiative and worked to develop a federated cloud architecture. In this hybrid model GEANT is now providing network connectivity to commercial data centres and EGI foresees the potential of the interoperation of its publicly funded sites with the commercial cloud providers.

Joint development with industry

On 1st February 2013 CERN hosted a workshop in the context of the CRISP project on the IT requirements for the next generation research infrastructures. More than 100 participants from the physics research infrastructure an IT industry participated to prepare a roadmap for future joint developments. A summary of the workshop is available online¹².

¹⁰ <https://cdsweb.cern.ch/record/1442597>

¹¹ www.helix-nebula.eu

¹² <http://indico.cern.ch/getFile.py/access?resId=0&materialId=0&confId=212402>

Relationship with other Activities

Research Data Alliance

The subjects addressed by the User Forum include data management aspects and hence are likely to be relevant to the recently formed Research Data Alliance¹³. The foreseen membership of the user forum includes leading research organisations that are the source of large and continuously growing data-sets actively used by the global research community. These represent '*big data factories*' that will be primary contributors of datasets to a future eInfrastructure commons. Hence the work and deliverables of the user forum will provide strategically important input for RDA working groups. The RDA governance model is still under evolution and it currently appears that the Technical Advisory Board would be a suitable interaction point.

eInfrastructure Reflection group (eIRG)

The infrastructure Reflection Group (eIRG) consists of official government delegates from all the EU countries. The e-IRG produces white papers, roadmaps and recommendations, and analyses the future foundations of the European Knowledge Society. The eIRG can be considered as orthogonal to the user forum which represents eInfrastructure users and research disciplines rather than countries. The results of the work of the User Forum will provide useful input to eIRG for its publications.

¹³ <http://rd-alliance.org/>

Process

In addition to the formal membership, there will be a steering group composed of a chairman and 3 other members that will:

- Be responsible for arranging the organization of the events. The intention is that each user community will be invited to host these events to encourage greater engagement from their community and network with representatives from other communities.
- Represent the results of the activities to the stakeholders and various international bodies.
- Fix the agenda and topics/issue to be discussed as a function of the interests of the members at large.
- Maintain the information repository and outreach utilities.

2-4 Meetings per year will be organized where working groups may be formed to address specific topics.

1-2 Open User Forums could be organized per year.

Working groups results will be hosted and made available on a website.

A mailing list of all members will be available.

Reports on the meetings of international bodies where the results of the forum are presented will be documented and made available through the website.

Next Steps

- Decide on an initial coordination team to bootstrap the activities.
- Approach relevant individuals to gauge the level of interest.
- Host an initial meeting to refine the terms of reference.
- Appoint a chairman and organizing panel.
- Disseminate the terms of reference and present the activity to relevant bodies.