

EGI.eu Position on Research Data Infrastructures: Framework for Action

Introduction

In preparation to H2020, the European Commission (DG CONNECT) has put together a  [Framework for Action](#) addressing the area of **Research Data e-Infrastructures**.

The “Action Fiches” propose concrete domains of action. They aim at realising the European vision for a global interoperable data e-infrastructure supporting open, digital-driven, science (note that High Performance Computing and Distributed Computing as such are outside the scope of this consultation).

By means of a public consultation the European Commission would like to get reactions on strengths, coverage, gaps etc in order to improve and fine tune the framework for action.

Period of consultation

The consultation will be on-line from 1 March until 27 March 2013.

To view the consultation document

 [Framework for Action](#)

Questionnaire for contributions

1. Please indicate the type of organisation you work for
(multiple answers possible)

- Research organisation
- Academia (university)
- Educational sector (e.g. teacher, student)
- Business
- Civil society (e.g. non-governmental organisation, trade union, consumer organisation)
- National public administration
- Regional/local administration
- EU public body
- International organisation
- Media

Other

(please explain)

2. Please indicate the key motivation to respond to the consultation

- You develop or operate data e-Infrastructures or an associated initiative
- You are establishing/operating Research Infrastructures
- You are a policy maker, a governmental official or work in a research funding agency dealing with research data
- You are a research practitioner
- Other, please specify below

Other

3. Please indicate your overall opinion on the "Framework for Action"

fully appropriate and timely

fully appropriate and timely, but some improvements are recommended (explain in section 4)

appropriate but substantial improvements are recommended (explain in section 4)

inadequate (explain in section 4)

4. Please provide your reactions to the "Framework for Actions" and its "action fiches"

Overall assessment of the relevance of the "Framework for Action"

The European Grid Infrastructure (EGI) federates compute, storage and data archives integrating them in a production pan-European infrastructure that provide access to more than 180 PB of data on disk and 160 PB on tape, supporting multiple disciplines. However, research communities request more and more access to multiple resource infrastructures (HTC for pre- and post-processing of data, HPC for simulation, permanent data storage etc). To achieve this goal a support action for the technical integration of computing and data infrastructure is needed for the provisioning of "common data services" (Fiche 05 Global Data Infrastructures). This action needs to be complemented by a support action that allows the integration of virtual research environments with these common data services (Fiche 01 Community support data services). EGI can play an important role by contributing to a global data infrastructure which is open and interoperable.

Identify strengths of one or more of the proposed fiches

Fiche 05 addresses the problem of infrastructure fragmentation, which is a current problem faced by users. A cross-cutting infrastructure of data services is important.

Fiche 01 community support data services: user community platforms need to be integrated in order to benefit from the common infrastructure of data services.

Identify gaps and propose changes to one or more of the proposed fiches

Fiche 05: global data infrastructures need to support the provisioning of software by Technology Providers. In addition the deployment of common data services should be pushed not only in data-centric infrastructures but also in other heterogeneous infrastructures like EGI, which already provide data storage, transfer, staging and indexing. Reuse of existing software should be promoted to avoid duplication.

Fiche 06: authentication and authorization, is not a specific problem of data infrastructure, needs to be addressed globally together with compute and network infrastructures to avoid fragmentation and lack of interoperability. Existing solutions for user authentication and authorization already in production should be reused. The development of a federated identity provisioning infrastructure shared by data and compute infrastructures should be facilitated.

Propose an additional fiche

We propose a cross-cutting fiche which addresses in an integrated way all of the following areas, ensuring that services and technical solutions mature into production-quality services:

- Support collaboration among multiple e-Infrastructures so that a single requirement gathering process involving user communities, infrastructure providers and technology providers is established
- Adaptation of virtual research environments to the common data service infrastructure
- Prototype a common data service infrastructure and support its evolution into a production service across multiple existing e-Infrastructures (EGI, PRACE, EUDAT, ESFRIs etc.).

5. Please provide us with background/references (a link to a report, a study, ...)

Please provide us with background/references

I would provide a link to:

The EGI process for requirements gathering, prioritization, and handling to technology providers - TCB Requirements Management document <https://documents.egi.eu/document/440>

EGI Use cases which show how data is analyzed and archived in EGI:

- The Pierre Auger Observatory http://www.egi.eu/news-and-media/newsletters/Inspired_Summer_2012/vo_auger_profile.html
- WLCG <http://www.egi.eu/cms/case-studies/WLCG.html>

First common EGI/EUDAT/PRACE workshop addressing the problem of infrastructure fragmentation and cross-infrastructure use case gathering: <https://indico.egi.eu/indico/conferenceTimeTable.py?confId=1228#20121126>