Guide to e-Infrastructure Requirements for European Research Infrastructures

An e-IRG support document

March 1, 2017

http://e-irg.eu

Colophon

Acknowledgments

The editorial board for this document
consisted of the e-IRG overarching working group
and was chaired by Arjen van Rijn and Rosette Vandenbroucke.
Many thanks go to all mentioned e-Infrastructures
who provided written material and additional input.

This document was produced by the e-IRGSP5
project for e-IRG.

e-IRGSP5 has received funding from the European
Union's Horizon 2020 research and innovation programme
under grant agreement No 730954

ISBN 978-90-823661-5-0
Persistent Identifier (e-IRG Catalogue) : http://e-irg.eu/catalogue/eirg-1004
Version 2.1

e-IRG secretariat

P.O. Box 93575,
NL-2509 AN The Hague,
The Netherlands
Phone: +31 (0)70 344 0526
E-mail: secretariat@e-irg .eu

Visiting address

Java Building, Laan van Nieuw Oost-Indië 300
NL-2593 CE The Hague
The Netherlands

Table of Contents

1 Executive Summary 5

2 Introduction 7

3 E-needs and guidance 9

3.1 Evaluation of proposals 9
3.2 Monitoring 2008 projects 11
3.3 Pilot periodic review Landmarks 13
4 The general e-Infrastructures landscape 15

4.1 Networking 15
4.2 Computing 16
4.3 Data infrastructures 17
4.4 Clouds 17
5 e-Infrastructure Commons and European Open Science Cloud 19

5.1 e-Infrastructure Commons 19
5.2 European Open Science Cloud (EOSC) 20
6 Access, FAIR principles and data management 21

6.1 Access to research infrastructures 21
6.2 FAIR Principles 22
6.3 Data Management 23
6.4 Other useful information about data infrastructures 23
7 A range of e-Infrastructure services for large-scale research infrastructures 25

7.1 Networking services 25
7.1.1 A separate internet – dedicated to research and education users 26
7.1.2 Management and Support 27
7.1.3 Trust, Identity & Security 28
7.1.4 Cloud Services 28
7.2 Computing services 29
7.2.1 PRACE Services 29
7.2.2 EGI services 30
7.2.3 HELIX NEBULA services 31
7.3 Data Services 32
7.3.1 EUDAT 32
7.3.2 OpenAIRE 34
8 Good practices and recommendations 35

9 References 36

10 Glossary 38

Foreword

Successful research and innovation needs access to first class research infrastructures and development of top-class research infrastructures is one of the key areas in the implementation of the European Research Area. In the overall context of research infrastructures, e-Infrastructures play a more and more important role. Today, almost all large-scale research activities include or are supported by several e-Infrastructure components. Major scientific breakthroughs are increasingly achieved by an international and cross-disciplinary team transferring, storing and analysing vast data collections and performing advanced simulations using different types of computing facilities. e-Infrastructures also enable a new mode of science where resources, in particular the other research infrastructures, are shared and used remotely to overcome fragmentation and cope with increasing costs and complexity.

To optimise the outcome for research and to facilitate the process of establishing a new research infrastructure, it is essential that e-Infrastructure needs are considered at the planning stages already. The European Strategy Forum on Research Infrastructures (ESFRI) is currently working on its Roadmap 2018[1], including a call for proposals for new Projects, the monitoring of the Projects listed on its Roadmap 2016 and a pilot periodic review of four Landmarks listed on its Roadmap 2016. Information on the needs for e-Infrastructures (e-needs) is requested in all these processes. Complementing the ESFRI Roadmap 2018 Guide[2], the main purpose of this document is to provide best practices for and guidance on the use of e-Infrastructures for large-scale pan-European research infrastructures in all phases of their lifecycle.

The current e-Infrastructure landscape in Europe can still be perceived as somewhat scattered but work is ongoing to remediate this inconvenience. More information can be found in the recently published e-IRG 2016 Roadmap[3]. This document also provides an overview of European e-Infrastructures and e-Infrastructure projects that are already providing services. This overview can be useful for identifying which European e-Infrastructure services are suitable for your research infrastructure and which organisation to contact to get more information.

Gabriele von Voigt

e-IRG Chair

March 2017

[1] http://www.esfri.eu/roadmap-2018

[2] http://www.esfri.eu/sites/default/files/docs/ESFRI_Roadmap_2018_Public_Guide_f.pdf

[3] http://e-irg.eu/catalogue/eirg-1002

1 Executive Summary

This e-IRG support document focusses on providing information and recommendations for complying with the e-needs instructions included in the ESFRI QUESTIONNAIRE FOR SUBMISSION OF PROPOSALS FOR ROADMAP 2018[4]. Information about existing e-Infrastructures (mainly at European level), about open science and about data management is included.

This document contains three main parts.

The first part addresses the e-needs instructions included in the ESFRI QUESTIONNAIRE FOR SUBMISSION OF PROPOSALS FOR ROADMAP 2018 and the corresponding key minimal requirements to be met.

The second part provides background information to be used when complying with the instructions and to follow up the recommendations and good practices presented in this document. We mainly describe the e-Infrastructures at European level, but are aware that e-Infrastructures at national or regional level are of equal importance. Efforts have been done to point to up-to-date information for e-Infrastructures at all levels. If needed, the e-IRG Secretariat and the national e-IRG Delegates can help you with contacting e-Infrastructures. This part also defines the European Open Science Cloud (EOSC)[5] and the e-Infrastructure Commons and contains information on access to research infrastructures and data management.

The third part discusses recommendations and best practices to be considered when complying with the e-needs instructions.

In summary the recommendations are:

- Check existing e-Infrastructures and related services for their use before defining the ICT infrastructure for your research infrastructure.
- Check with existing research infrastructures how they realised their ICT infrastructure.
- Contact existing e-Infrastructures at national and European level and ESFRI Projects and Landmarks - as appropriate.
- Work to an ICT synergy with other Projects and Landmarks to encourage the development of the e-Infrastructure commons.

[4] http://www.esfri.eu/sites/default/files/docs/ESFRI%20Roadmap%202018_Proposal_Submission_Questionnaire_Public.pdf

[5] See Chapter 5

- Pay attention to the interoperability of services and data.
- Plan your access to ICT resources.
- Take into account the FAIR principles for your research data.
- Define a data management plan that fits your Research Infrastructure.
- Pay attention to the interoperability of services and data.
- Plan your access to ICT resources.
- Take into account the FAIR[6] principles for your research data.
- Define a data management plan that fits your Research Infrastructure.

[6]See Chapter 6

2 Introduction

The aim of this support document is to provide guidance to large-scale research infrastructures in the planning and implementation of their electronic infrastructure (e-Infrastructure) components, with a special focus on integration with and use of existing or emerging national and European electronic Infrastructures (e-Infrastructures). Most of the large-scale international research infrastructures have significant networking, computing and data management needs. In the past, such research infrastructures often established their "own" e-Infrastructure components and implemented and operated the related services. This was done to ensure that the best possible services, adapted to the particular research infrastructure environment and users' requirements, were provided. However, both national and European e-Infrastructures of more generic nature have evolved significantly and now offer a broad range of services developed in collaborative efforts with users. It might be more cost effective and efficient to use generic e-Infrastructure services than to build a research specific e-Infrastructure.

Existing and emerging e-Infrastructures represent a highly valuable asset when planning and implementing large-scale research infrastructures and such infrastructures in turn represent important user communities for more generic e-Infrastructures. Much effort has been spent on bringing research infrastructures and e-Infrastructures together for their mutual benefit, but it might still be difficult for the stakeholders to grasp the needs of users or the advantages of using existing e-Infrastructures and related services.

The call for proposals for the ESFRI Roadmap 2016 for the first time included questions related to the e-needs of the proposed research infrastructures. The ESFRI Strategy Working Groups (SWGs) evaluated the answers on those questions. The experiences have been reported[7] and have led to an updated formulation of the instructions on e-needs. Those new instructions are integrated in the call for proposals for the ESFRI Roadmap 2018 update. Furthermore, the e-needs instructions have been formulated in the specific questionnaires for the monitoring of the 2008 & 2010 ESFRI Projects and for the pilot periodic review of four ESFRI Landmarks. This differentiation reflects the changing needs of a research infrastructure during its progress towards implementation and its lifecycle. All instructions are listed in chapter 3 accompanied by a list of the related key minimal requirements in the different phases of the lifecycle with corresponding explanations. These minimal key requirements have to be met by the new proposals as well as by the listed Projects and Landmarks.

[7] http://e-irg.eu/documents/10920/239416/Experiences+of+e-IRG+involvement+in+ESFRI+proposal+evaluation.pdf

To be able to comply with the e-needs instructions, it is important to have an understanding of the e-Infrastructure landscape at the regional, national and European levels. Hence, we provide a short overview of the basic e-Infrastructure components focussed on the European scale. However, note that in many cases also national or regional e-Infrastructures can be useful to ESFRI initiatives and that often coordination exists between the national/regional and European level. It is not possible to elaborate all details in this document on the national e-Infrastructures, but instead pointers are provided in the form of references in the e-IRG Knowledge Base[8]. In general, contact information and pointers to more information about e-Infrastructures are presented in the References section. Also, the implementation of the FAIR principles requires to assess and to take into account the international and European disciplinary data landscape.

[8] http://knowledgebase.e-irg.eu

3 E-needs and guidance

3.1 Evaluation of proposals

Proposals need to comply with the following key minimal requirements for e-needs for the preparation phase:

- conceptual design of e-Infrastructure ready;
- contributions of e-Infrastructure resources at all levels (institutional, regional, national, international) described;
- access policy and Data Management Plan (DMP) outlined.

Nr	Instruction	Guidance
4.1	Outline the Data Management Plan (DMP) and data access policy of the RI. If applicable, describe how data would become accessible to the public (maximum 3000 characters with spacing)	Explicit data policy with justification, including access rights and restrictions, policy for long-term usability, policies for data formats and meta-data structures, policy for availability within, or interoperability, with existing data sharing frameworks. Procedures and resources needed for data collection, documentation, storage and access, long-term preservation, including usability, availability, time-frame. Application of FAIR principles. Draft agreements involving future users (e.g. embargo period

		and IPR issues). Answer might refer to existing and documented policies.
4.2	Describe and quantify what e-Infrastructure services - e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - your RI will need: (maximum 2000 characters with spacing)	The RI should quantify in some order of magnitude its ICT-needs in commonly used units for computing (core hours, SpecInts, Flops), data storage (TB, PB), networking (Mbps, Gbps). The RI should also describe its needs in terms of required services, such as Authentication and Authorization Infrastructure, compute services (e.g. HPC, HTC, cloud), data services, network services, etc. If the RI cannot describe this, it should be explained why not.
4.3	Describe how the e-Infrastructure services needed by your RI will be implemented, specifying the potential need of external e-Infrastructure resources and the relations to external e-Infrastructures: (maximum 2000 characters with spacing)	Based on the answer on 4.2 the RI should describe, how it plans to meet the requirements, e.g. implemented as part of the RI infrastructure construction cost or relying on external public (institutional, regional, national, international) or private service providers or a combination of internal and external provisioning.
4.4	Describe how the RI will contribute to the development of the European and global e-Infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-Infrastructure Commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)	The RI should demonstrate knowledge about the key components of an e-Infrastructure commons or EOSC (terms maybe used interchangeably), in particular on the interface between generic and RI specific e-Infrastructures. They should make clear how their DMP and data access policy fit in the objectives of the EOSC. The RI should also explain how it contributes tot he development of the disciplinary and generic data landscape.

3.2 Monitoring 2008 projects

2008 Projects need to comply with at least the following key minimal requirements for e-needs for the implementation phase:

- technical design of e-Infrastructure ready and approved;
- draft operational planning for e-Infrastructure service delivery;
- agreements with parties delivering core e-Infrastructure services (Central Hub in case of distributed research infrastructures) drafted;
- access policy and DMP approved, including plan for sustainability of data;
- security policy defined and approved.

Nr	Instruction	Guidance
4.1	Summarise the data management, data access and data security policies of your RI – including procedures for ensuring the sustainability of data – and clarify to what degree they are approved (maximum 3000 characters with spacing)	Explicit data policy with justification, including access rights and restrictions, policy for long-term usability, policies for data formats and meta-data structures, policy for availability within, or interoperability, with existing data sharing frameworks. Procedures and resources needed for data collection, documentation, storage and access, long-term preservation, including usability, availability, time-frame. Application of FAIR principles. Draft agreements with users on data management, embargo period and IPR. Answer might refer to existing and documented policies (see 4.5).
4.2	Describe the technical design and operations of the e-Infrastructure services - e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - for your RI and clarify to what degree it is approved *and implemented*: (maximum 2000 characters with spacing)	The answer should outline the architecture of the system that addresses the ICT-needs. It should contain a quantification of the ICT-needs in commonly used units for computing (core hours, SpecInts, Flops), data storage (TB, PB), networking (Mbps, Gbps) and a description of the defined services, such as Authentication and Authorization Infrastructure, compute services (e.g. HPC, HTC, cloud), data services, network services, etc. The approval and implementation status of the system design should be described.
4.3	Describe the use of external e-Infrastructure resources and the relations to external e-Infrastructures - including	If the technical design described in 4.2 relies on the use of external e-Infrastructure resources, this relation should be described here. External resources or services can be public (institutional, regional, national, international), private (commercial) or a combination. The status of agreements

	agreements with external partners delivering core e-Infrastructure services for your RI - and clarify to what degree this use is approved within your consortium: (maximum 2000 characters with spacing)	for resource and service provisioning should be included in the description.
4.4	Describe how the RI contributes to the development of the European and global e-Infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-Infrastructure Commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)	The RI should demonstrate knowledge about the key components of an e-Infrastructure commons or EOSC (terms maybe used interchangeably), in particular on the interface between generic and RI specific e-Infrastructures. It should make clear how its DMP and data access policy fits in the objectives of the EOSC. The RI should also explain how it contributes to the development of the disciplinary and generic data landscape.
4.5	If available, upload the approved Data Management Plan (DMP): (upload with limit 1 MB)	2008 Projects must provide an approved DMP.

3.3 Pilot periodic review Landmarks

Landmarks need to comply with at least the following key minimal requirements for e-needs for the implementation phase:

- technical design of e-Infrastructure ready and approved;
- draft operational planning for e-Infrastructure service delivery;
- agreements with parties delivering core e-Infrastructure services (Central Hub in case of distributed research infrastructures) drafted;
- access policy and DMP approved, including plan for sustainability of data;
- security policy defined and approved.

Landmarks in the operation phase, need to comply with the following key minimal requirements for e-needs:

- operational planning ready and approved;
- agreements with service provisioning parties signed;
- DMP implemented and security policy deployed.

Nr	Instruction	Guidance
4.1	Summarise the data management, data access and data security policies of your RI – including procedures for ensuring the sustainability of data – and clarify to what degree they are approved *and implemented* (maximum 3000 characters with spacing)	Explicit data policy with justification, including access rights and restrictions, policy for long-term usability, policies for data formats and meta-data structures, policy for availability within, or interoperability, with existing data sharing frameworks. Procedures and resources used for data collection, documentation, storage and access, long-term preservation, including usability, availability, time-frame. Application of FAIR principles. Agreements with users on data management, embargo period and IPR. Answer might refer to existing and documented policies (see 4.5).
4.2	Describe the technical design and operations of the e-Infrastructure services -e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - for your RI and clarify to what degree it is approved *and implemented*: (maximum 2000 characters with spacing)	The answer should outline the architecture of the system that addresses the ICT-needs. It should contain a quantification of the ICT-needs in commonly used units for computing (core hours, SpecInts, Flops), data storage (TB, PB), networking (Mbps, Gbps) and a description of the used services, such as Authentication and Authorization Infrastructure, compute services (e.g. HPC, HTC, cloud), data services, network services, etc. The approval and implementation status of the system design should be described.

4.3	Describe the use of external e-Infrastructure resources and the relations to external e-Infrastructures - including agreements with external partners delivering core e-Infrastructure services for your RI - and clarify to what degree you have *signed agreements*:: (maximum 2000 characters with spacing)	If the technical design described in 4.2 relies on the use of external e-Infrastructure resources, this relation should be described here. External resources and services can be public (institutional, regional, national, international), private (commercial) or a combination. The status of agreements for resource and service provisioning should be included in the description.
4.4	Describe how the RI contributes to the development of the European and global e-Infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-Infrastructure Commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)	The RI should demonstrate knowledge about the key components of an e-Infrastructure commons or EOSC (terms maybe used interchangeably), in particular on the interface between generic and RI specific e-Infrastructures. It should make clear how its DMP and data access policy fits in the objectives of the EOSC.
4.5	If available, upload the approved Data Management Plan (DMP): (upload with limit 1 MB)	Landmarks must provide a DMP.

4 The general e-Infrastructures landscape

The pan-European e-Infrastructures for networking, high-performance computing (supercomputing) and high-throughput computing (clusters built from more commodity-type hardware) are already well established and provide production services used by international research and research infrastructures. Also, data and cloud infrastructures are developing fast.

It should be noted that the European-level e-Infrastructure services are often provided by national e-Infrastructures in a collaborative setting and that the European initiatives are dependent on the existence of strong and coherent national e-Infrastructure nodes and of their cooperation to enable cross-border services for scientific communities.

Available e-Infrastructures and their related services might not always fulfil the user's needs. But collaboration and coordination between e-Infrastructures and between e-Infrastructures and research infrastructures is ongoing.

The following paragraphs present a brief introduction of the major pan-European horizontal e-Infrastructure initiatives. An overview of the available services is provided in Section 7.

4.1 Networking

Connecting research communities across the globe is a prerequisite to stimulate exchange of ideas, data and results. Moreover, it is needed to provide remote access to unique research facilities and to shared data that are located at specific places; in this way, the quest for breakthroughs is sped up, as large communities can simultaneously be part of joint investigations and separate research teams can exploit large infrastructures and data in parallel.

Already since a few decades the National Research and Education Networks (NRENs) have been connecting universities, research institutes, and sometimes other public institutions in their country. Their governance and sources of income differs from country to country, as well as the way the access to NREN resources is managed. The GÉANT Association has gradually grown into a pan-European organisation, where the associated NRENs link together research communities and provide trans-national access to research infrastructures and research resources. GÉANT provides interconnectivity between NRENs across 43 European countries, by serving an estimated 50 millions of users of practically all research disciplines and thematic domains. The major scientific communities that are connected via GÉANT's federated network are amongst others high-energy physics, bio-

medical sciences, astronomy/radio astronomy, earth observation and early warning, as well as arts and culture.

In addition to pan-European connectivity, the GÉANT network has international connections to a large set of partner networks (some 60 NRENs) worldwide, in particular through regional agreements – thereby enabling international collaboration for research and education.

Most large-scale research infrastructures can connect to the local NREN and thus access GÉANT, enabling worldwide communications. Projects can also work with their related NRENs and GÉANT for international *point-to-point links* to connect parts of the research infrastructure that are distributed over Europe or beyond. If the project or infrastructure is distributed across national boundaries, GÉANT can *help coordinate with the relevant local NRENs* and *advise on appropriate technical solutions*. GÉANT provides also important experimental services for researchers, such as *innovation test beds.*

4.2 Computing

The role of computing in science has always been prominent, especially in those areas where models are simulated to predict and explain observed phenomena. Recently computing applied to (big) data has transformed our way of doing science, and it has further stimulated the need of computing resources. These needs range from powerful PCs via clusters to high-performance computers.

On a national level the varieties of high throughput computing resources are often brought together using a *national grid infrastructure (NGI)*. The NGI's are federated in a pan-European high-throughput computing infrastructure (European Grid Infrastructure, or EGI). EGI also features a federated cloud infrastructure based on the NGI's cloud infrastructures. The High Performance Computing (HPC) national infrastructures are federated in the Partnership for Advanced Computing in Europe (PRACE). EGI focuses on large-scale federated high-throughput and cloud computing solutions, while PRACE offers access to world-class high-performance capability computing facilities and services. EGI and PRACE are respectively managed by the organisations EGI.eu and PRACE AISBL.

Both EGI and PRACE have established contacts with consortia that operate or prepare European large-scale research infrastructures to understand their needs and to find out how these match with available resources and existing policies.

PRACE systems are available to scientists and researchers from academia and industry from around the world through the process of submitting computing project proposals based on " Excellence of science" and supported by scientific peer-review. There are basically two forms of access:

- *preparatory* access, intended for short-term access to resources, for code-enabling and porting, required to prepare proposals in the category "project access" and to demonstrate the scalability of codes;
- *project access,* intended for individual researchers and research groups including multi-national research groups, which can be used for 1-year, 2-year or 3-year (Multi-Year Access) production runs.

EGI provides solution frameworks built through a service catalogue that has been evolving during many years. The *EGI Federated Cloud Solution* offers a standards-based and open infrastructure to deploy on-demand IT services that can host datasets of public or commercial relevance, and can be flexibly expanded by integrating new providers. This is complemented by the *EGI High Throughput Computing Solution* that provides a global high-throughput data analysis infrastructure, linking a

large number of independent organisations and delivering computing resources with high scalability. The *EGI Federated Operations Solution* and the *EGI Community-Driven Innovation & Support Solution* are provided for management of a heterogeneous infrastructure and are helping researchers and research infrastructures to access and to use computational services.

Access to EGIs externally provided resources and to national HPC resources, is provided through various access modes, such as free grant-based allocations, pay per use, and annual membership fees. For the national HPC infrastructure(s), the access modes are closely connected to the ruling national governance.

4.3 Data infrastructures

Data is a key research infrastructure. This was stressed in the "Riding the Wave" report of the relevant EC High Level Expert Group in 2010 and its follow-up RDA Europe Report "The data Harvest" (2014), as well as in the recent Communication from the European Commission "European Cloud Initiative - Building a competitive data and knowledge economy in Europe". Science is data-driven. As explained in the strong statement of the G8 Ministers of Research in 2013[9]:

'To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards."

Data has to be open (except for legitimate restrictions such as privacy), FAIR - "Findable/Accessible/Interoperable/Reusable" - and preserved on the long term. The users of the data infrastructures and services are the data providers and data consumers, who can belong to the scientific community, to industry, to the public sector or can be citizens.

Data infrastructures and data services need to be part of the e-Infrastructure Commons (see Chapter 5). Data infrastructures should be built in an interoperable way and provide all potential users with the capability to store their data and to make this data discoverable and accessible while taking into account European and national data laws (privacy, IPR, …).

The way to such an interoperable European Data Infrastructure might be long. Initiatives at the European level have been started to store data of researchers in general (EUDAT) or to store research publications (OpenAIRE). Scientific communities and research infrastructures have been building their framework for data sharing (formats, metadata …) and in many cases their own data infrastructures, often at the international level, taking into account their specific needs. National, regional and local authorities also set up data infrastructures. All of them should be interconnected in a European Data Infrastructure, which should be an ecosystem able to include different components.

4.4 Clouds

Cloud technologies have well made their entrance in non- research environments where probably the most visible use is "Software as a Service". In the research environment local/national cloud services have gained some success but at the European level the cloud initiatives are scarce and

[9] http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf

have not led to a pan-European cloud e-Infrastructure. Initiatives to mention include the *Helix Nebula initiative,* the *EGI federated cloud* and the *GÉANT cloud services platform* in this field**.** Helix Nebula is providing a channel by which innovative cloud service companies can work with major IT companies and public research organisations. The Helix Nebula Marketplace (HNX) is the first multi-vendor product coming out of the initiative and delivers easy and large-scale access to a range of commercial Cloud Services through the innovative open source broker technology. Also, GÉANT is actively helping NRENs (National Research and Education Networks) to deliver cloud services to their communities. It is also engaging with the existing NREN brokerages to promote an efficient and coordinated pan-European approach, by building on existing experience and supplier relationships. Recent announcements of the European Commission (European Open Science Cloud) and the Cloud Communication have added new pressure for the adoption of cloud technologies and services in the research environment although the content of these announcements can also been seen as a metaphor of a federation of seamlessly accessible resources and services including commercial services in order to provide a more complete offer to researchers.

5 e-Infrastructure Commons and European Open Science Cloud

5.1 e-Infrastructure Commons

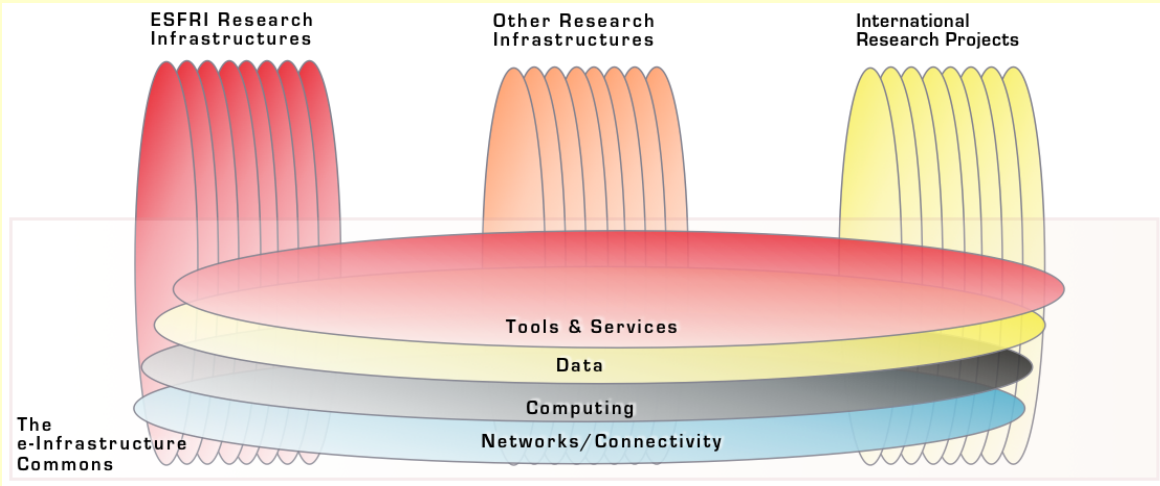

Illustration 1: The e-Infrastructure Commons provides a common interface for research infrastructures to access e-Infrastructures.

Already in 2012, e-IRG pointed to the need of *a single e-Infrastructure Commons for knowledge, innovation and science as a living ecosystem that is open and accessible and continuously adapts to the changing requirements of research*. The notion of an e-Infrastructure Commons has now been widely accepted, and significant steps have been taken towards its implementation. The e-Infrastructure Commons is the framework for an easy and cost-effective shared use of distributed electronic resources for research and innovation across Europe and beyond. An essential feature of the Commons is the provisioning of a clearly defined, comprehensive, interoperable and sustained set of services, provisioned by several e-Infrastructure providers, both public and commercial, to fulfil specific needs of the users. This set should be constantly evolving to adapt to changing user needs and new technological capabilities, complete in the sense that the needs

of all relevant user communities are served and minimal in the sense that all services are explicitly motivated by user needs and that any overlap of services are thoroughly justified.
Based on this description, the three distinct functions of the Commons can be summarized as follows:

- A platform for *coordination* of the services building the Commons, with a central role for (European) research, innovation and research infrastructure communities.
- *Provisioning* of sustainable and interoperable e-Infrastructure services within the Commons, promoting a flexible and open approach where user communities are empowered to select the services that fulfil their requirements.
- Implementation of innovation projects aiming for the constant *evolution* of e-Infrastructures needed to meet the rapidly evolving needs of user communities and take advantage of relevant new technologies when they are mature enough.

The above ideas have contributed to the establishment of European initiative now known as the European Open Science Cloud (EOSC), discussed in the next paragraph. In practice both terms, EOSC and e-Infrastructure Commons can be used interchangeably.

5.2 European Open Science Cloud (EOSC)

With the adoption of the Digital Single Market's strategy in May 2015, the European Commission announced the launch of a cloud for research data – the 'research open science cloud'. In conjunction with the Digital Single Market (DSM)[10], the declaration of a "European Open Science Cloud" (EOSC)[11] followed. The EOSC aims to create a trusted environment for hosting and processing research data to support EU science by offering 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with free access at the point of use, open and seamless services for storage, management, analysis and re-use of the data that are linked to their research activities, across borders and scientific disciplines. The European Council, together with the Member States and the European Parliament welcomed the initiative.

Moreover, the European Parliament also called upon on the European Commission, in cooperation with all relevant stakeholders, to set up an action plan to lead to the establishment of the European Open Science Cloud, which should seamlessly integrate existing networks, data and high- performance computing systems and e-Infrastructure services across scientific fields, within a framework of shared policies, standards and investments. A High Level Expert Group on the European Open Science Cloud (HLEG EOSC) was established and its first report has been published in October 2016[12]. A pilot project started in January 2017.

[10] https://ec.europa.eu/digital-single-market/cloud
[11] http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
[12] http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none

6 Access, FAIR principles and data management

The currently growing amount of information on the Internet requires a common data infrastructure in parallel to the existing networking and computing environments.

e-InfrastructuresResearch infrastructures, such as the Projects and Landmarks listed on the ESFRI Roadmap, produce and are dependent on a rapidly increasing amount of data. For research and society to take full benefit of the major investments in research infrastructures, the data needs to be made openly and easily available for researchers, over wide spans of fields, in sustainable settings. To enable this, the data needs to be managed, stored and preserved in a cost-efficient way. The access to the data across borders and domain boundaries must be secured. Some existing e-Infrastructures provide versatile services and tools needed for both data management and access. The e-needs instructions - in particular 4.1 - address access to research infrastructures and data. They request to respect the FAIR principles (see below) and to provide a data management plan.

6.1 Access to research infrastructures

The European Commission issued a "European Charter for Access to Research Infrastructures"[13] with input from all relevant stakeholders.

The Charter follows the July 2012 ERA Communication on "a Reinforced European Research Area Partnership for Excellence and Growth" and has the purpose of setting out non-regulatory principles and guidelines to be used, on a voluntary basis, as a reference when defining or re-defining rules and conditions for access to research infrastructures.

In such respect, the Charter promotes the harmonisation of Access procedures as well as the enhanced transparency of access policies adopted by research infrastructures with the final purpose of enabling Users to Access the best research infrastructure to perform their work, wherever it might be located.

The Charter is primarily targeted at those responsible for the definition of access rules and conditions to any given Research Infrastructure and, therefore, at the Research Infrastructures

[13] http://ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf#view=fit&pagemode=none

themselves, at the institutions to which they belong and at their respective research funding organisations.

The Charter is applicable to any type of research infrastructure, including the e-Infrastructure dimension. It also points to Reference Documentation where more information to data can be found[14].

6.2 FAIR Principles

The e-needs instruction 4.1 request to apply the FAIR+R principles where FAIR stands for Findable, Accessible, Interoperable and Re-usable. Sometimes +R is added which stands for 'Reproducible. Force11[15] provides a set of guiding principles to make dataFAIR.

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. Metadata specify the data identifier.

To be Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

To be Re-usable:

R1 meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

More detailed information can be found in "Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing" version b 1.0"[16]

Research infrastructure participation in international initiatives set up to define the disciplinary interoperability framework is important to make sure that the framework fits their needs and to make the implementation of its recommendations easier.

[14] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=access_ri
[15] https://www.force11.org/about
[16] Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0

6.3 Data Management

Data infrastructures built by research infrastructures or other players in the research field are often customised for the concerned project or research discipline domain at the best and do not allow for use beyond the project or discipline borders. Interoperability between data infrastructures at all levels (storage, metadata, services) becomes crucial. These facts have been confirmed by many projects and researchers, and efforts have been made to attain a common understanding on the realisation of an ecosystem of data infrastructures and related services. As explained above, a component of interoperability is at the disciplinary level, to define and document common formats and metadata allowing data discovery, exchange and re-use beyond the borders of a specific project. Many disciplines work at the European and international level to define the discipline-specific aspects of their data infrastructure, which have to be interfaced with the generic components of the data infrastructure, such as the ones provided by EUDAT as described below. The Research Data Alliance (RDA) hosts some of these discussions.

Including a data management policy already at the planning stage is vital for large scale research infrastructures and research projects. Each project/infrastructure proposal should also have a data management plan. ESFRI Projects and Landmarks - where the users (research projects) generate the data - need to clarify which entities should actually develop the data management plan and who is in charge of its implementation. The data management policy addresses access rights and restrictions, long-term usability, data formats, meta-data structures, availability within, or interoperability, with existing data sharing frameworks. The data management plan outlines how data will be handled during the research and after completion of the research project. A data management plan can include data services and plans for collaboration between projects at the data level. It also specifies procedures and resources needed for data collection, documentation, storage and access. It can reference procedures and resources needed for long-term preservation, including usability, availability, time-frame, etc. associated with a funding plan. The data management plan should be consistent with the data management policy of the research infrastructure.

There is no general data management plan template available that will suit all research domains and each research infrastructure. Examples of data management plans are available and mentioned in the References section.

6.4 Other useful information about data infrastructures

In this complex world of data and data management it is relevant to check the work done by the ESFRI Cluster projects, RDA and e-IRG itself. This information can provide a base for the data management in the future research infrastructures. The ESFRI Cluster projects BioMedBridges, ENVRI, CRISP and DASISH have published the document "Realising the full potential of research data: common challenges in data management, sharing and integration across scientific disciplines"[17]. The document tackles among others common data standards and formats, data storage facilities, integrated access and discovery, data curation, privacy and security, service discovery and service market places. Other Clusters such as ASTERICS (astronomy and astroparticles) work on the inclusion of ESFRI data in the international disciplinary interoperability framework and liaison with the RDA.

RDA[18] hosts a large palette of activities aimed at facilitating scientific data sharing, including the definition of the disciplinary interoperability framework for disciplines which do not have a specific organisation for that purpose. They also address community needs, data stewardship and sharing,

[17] http://zenodo.org/record/7636

[18] http://www.rd-alliance.org

including data publication and citation, and more technological topics related to the base infrastructure of data sharing. Its growing set of recommendations and output[19] covers sociological and technological aspects of data sharing.

Additional work topics can be proposed by any RDA individual member, and if interest from participants from different regions is identified they can be proposed as a new Interest Group, or if implementable deliverables are foreseen within 18 months, as a new Working Group. The RDA Europe project recently set up the Group of European Data Experts[20] (GEDE) to gather needs and feedback from the ESFRIs and other large European projects and to build consensus on the needed building blocks and guidelines.

RDA Europe published in 2014 its report "The Data Harvest: How sharing research data can yield knowledge, jobs and growth"[21] in which they call for the sharing of data over vast distances, across institutions and disciplines.

[19] https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs

[20] https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda

[21] https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf

7 or

e-Ir ent sections of this chapter. The table below gives a quick overview of e-Infrastructure services and their providers and which access mode is used by those providers.

e-Infrastructure	GÉANT	EGI	PRACE	HELIX NEBULA	EUDAT	ZENODO	OpenAIRE
Services							
Network	x						
HTC		x					
HPC			x				
Cloud	x	x		x			
Data		x			x	x	x
Access Modes							
excellence driven		x	x				
market driven	x	x		x	x		
wide access	x	x			x	x	x

7.1 Networking services

A range of networking services is delivered by GÉANT at the international level. Most of these services match those offered at national level by the NRENs. The GÉANT Network Compendium (references in Annex 1) gives an overview of all NRENs, their services and the contact information. The GÉANT services are described below.

GÉANT offers a wide range of innovative services to enhance the experience of its network. GÉANT's portfolio comprises a range of aligned services that allow it to cater for the most varied and demanding needs.

7.1.1 A separate internet – dedicated to research and education users

The GÉANT and NREN networks underpin the work of a wide range of e-Infrastructure and scientific research projects by providing a high performance, reliable and cost-effective communications platform across the research and education (R&E) community. Service options cover IP, dedicated private connections, virtual private networks and roaming options.

IP Networking

The core of the GÉANT network is the world-class IP backbone. GÉANT IP provides general-purpose IP transit for national research and education networking (NREN) organisations and other approved research and education partners and providers.

Its function is to provide a private service for IP (internet protocol) traffic thatis separated from general-purpose access to the internet. Offering users connection speeds of up to 100Gbps, GÉANT IP provides the essential communication service that supports inter-NREN connectivity..

VPN Services

Many projects may require teams across Europe to be able to collaborate effectively with enhanced privacy. By creating a Virtual Private Network (VPN), all sites on the VPN can communicate without the need to arrange separate physical networks, while benefiting from the privacy and security of a private infrastructure. GÉANT can provide VPNs between many sites over great distances within Europe and reach the USA (via Internet2 and ESnet), Canada (via CANARIE), and Asia.

GÉANT offers two core VPN services:

Layer 3 VPN - The GÉANT L3-VPN service provides NRENs with the backbone infrastructure to enable custom VPN services for their users across the GÉANT backbone.

GÉANT Multi-Domain Virtual Private Network (MD-VPN) - MD-VPN provides an end-to-end international network service that enables scientists all over Europe to collaborate via a common private network infrastructure.

In addition to the core IP and VPN networking services, GÉANT offers users a range of specialised connectivity options. Many performance-critical services require guaranteed performance levels and additional security that is difficult to achieve through shared IP services. In particular, applications such as data centre backup and replication, real-time mission-critical services and broadcast quality video need the guaranteed bandwidth and low latency that only dedicated circuits separated from general IP traffic can offer.

Point-to-point services

Point-to-point services provide dedicated connectivity between two sites over the existing infrastructure without the cost and difficulty of building and managing a dedicated physical network. This type of connectivity can provide fixed latency between collaborating institutions, a high level of security and, if needed, guaranteed bandwidth of up to 100Gbps. GÉANT offers a range of point-to-point services.

- GÉANT Plus - The GÉANT Plus service allows national research and education networking (NREN) organisations to request point-to-point Ethernet circuits between end-points at GÉANT PoPs (points of presence). Circuits can be established to any European NREN.
- GÉANT Lambda - GÉANT Lambda provides full wavelengths at up to 100Gbps to support NREN users with particularly demanding network requirements. The service provides transparent 10Gbps, or 100Gbps wavelengths between GÉANT PoPs. GÉANT Lambda services can be provided unprotected or restored using GMPLS (Generalised Multiprotocol Label Switching) signalling.
- GÉANT bandwidth on demand - GÉANT Bandwidth on Demand (BoD) is a flexible network option for situations where users need to reliably transfer large data sets for short periods of time between two end points, with guaranteed bandwidth. The service is offered collaboratively by GÉANT and participating European NRENs, delivering a seamless user experience from order to delivery.

Open Interconnectivity

As international and public-private partnerships grow in importance within the R&E sector, so the high-performance, flexible and neutral interconnection points provided by the GÉANT Open service can offer new opportunities. Users can connect their own circuits – at 1Gbps, 10Gbps or 100Gbps – and can then request interconnections with any other participant.

GÉANT Testbeds Service

The GÉANT Testbeds Service (GTS) delivers integrated virtual environments as 'testbeds' for the network research community. GTS is designed for researchers of advanced networking technologies to help support testing and development over a large-scale, dispersed environment. GTS can support multiple projects simultaneously and isolates them from each other and from the production GÉANT network to provide security and safety. This facility is leading the way in providing facilities to help develop the next generation of internet services.

eduroam

provides 50 million students and researchers with access to thousands of wi-fi access points in over 70 countries using a single, secure login facility - making international collaboration much easier. Over 5 million international logins a day are enabled by eduroam.

7.1.2 Management and Support

The connectivity delivered by GÉANT is supported by a comprehensive range of network monitoring and management services. They optimise network performance by providing 24x7 monitoring across the GÉANT Service Area infrastructure, enabling fast identification and remedy of any faults on the network as well as providing powerful security to prevent and detect malicious attacks.

Users benefit from the range of GÉANT network monitoring, security and support services employed by NRENs to assure optimum performance for projects and institutions. The areas of tools and services in this group include performance measuring and monitoring, performance enhancement and security.

- Performance measuring and monitoring - Analysing performance in global research networks is complex since any single path might go through several domains – campus, local and national networks as well as the GÉANT backbone. Offering comprehensive multi-domain monitoring features, GÉANT's perfSONAR services allow users to access network performance metrics and perform network monitoring actions across multiple domains, ensuring that any source of congestion or outage on a point-to-point connection can be quickly and easily identified and addressed.
- Performance enhancement - The Performance Enhancement Response Team (PERT) provides an investigation and consulting service to academic and research users on their network performance issues. The service is achieved via eduPERT, the federated structure that combines the PERTs from the local institutions, NRENs and GÉANT and fosters knowledge-sharing across the GÉANT network community. eduPERT is part of GÉANT's commitment to helping users get the best performance from their connections.
- Security - In an online world, network security is of paramount importance. GÉANT takes a proactive approach to security to maintain the integrity of the network, implementing advanced defences that offer sophisticated handling of network incidents.

By providing strategies for incident prevention, detection and handling, the GÉANT security systems will allow users to keep network domains secure by monitoring traffic and routing information.

7.1.3 Trust, Identity & Security

GÉANT and its NREN partners provide technologies that build trust, promote security and support the use of online identities. This is an essential component of many infrastructure projects by bringing together services and users in a scalable, manageable and secure manner.

EduPKI - Digital certificates are issued by Certification Authorities (CAs) and are widely used to guarantee secure and reliable communication between servers, users, or between a user and a server. The eduPKI service aims to assist the adoption of digital certificates within the project in a cost-effective way. It aims to create a service able to support other GÉANT services in defining their security requirements, and to provide them with digital certificates. eduPKI builds on top of existing NREN CA services, federating them to make all participating CAs available to GÉANT services.

eduGAIN enables the trustworthy and secure exchange of authentication, authorisation and identity (AAI) information. It interconnects identity federations around the world, simplifying access to content, services and resources. eduGAIN provides a pan-European Web Single Sign On (Web SSO) (i.e. a single digital identity and password) to access all services provided by the participating federations and their affiliated service providers. This service is of special interest for distributed infrastructures or data archives, allowing data to be retained locally while researchers access data sets from different locations via a single sign on.

7.1.4 Cloud Services

Cloud services offer higher education and research organisations the opportunity to become more agile and provide their users with a wider range of IT services at a lower cost. GÉANT provides the platform for users to access cloud services and, through its cloud service catalogue, works with other e-Infrastructure projects and commercial cloud service providers to help deliver innovative services to research and education institutions and their users.

GÉANT is actively helping national research and education networking organisations (NRENs) to deliver cloud services to their communities, with the right conditions of use. For example GÉANT and the NRENs have made arrangements with two of the biggest cloud service providers, Amazon and Microsoft, to dramatically reduce the data connectivity costs associated with cloud computing - providing real cost savings to the R&E community. The goal is an attractive, well-balanced portfolio of cloud services to support European research and education.

7.2 Computing services

7.2.1 PRACE Services

PRACE provides HPC time to researchers in Europe on a set of seven Tier-0 systems provided by 5 Hosting Members (France, Germany, Italy, Spain and Switzerland). In addition PRACE also offers high value services like the High Level Support Teams (HLSTs) in charge of supporting European scientific communities in their efficient use of Tier-0 systems, extended training program through six PRACE Advanced Training Centers, code porting/optimization activities with the support of PRACE experts and support to industry to enable the adoption of HPC by industry, including SMEs. The European countries that are member of PRACE also offer HPC services to the local research community and often to industry and likewise offer relevant courses to the researchers.

PRACE HPC access

PRACE systems are available to scientists and researchers from academia and industry from around the world through the process of submitting computing project proposals based on open R&D. When

granted on the sole criteria of scientific excellence, research projects can have access for free to PRACE HPC resources for a duration between one and three years with the condition to publish results. Two project proposal models are now in use: those for preparatory access and those for full project access. Project Access is subject to the PRACE Peer Review Process, which includes technical and independent scientific review. Technical experts and leading scientists evaluate the proposals submitted in response to the bi-annual calls. Applications for Preparatory Access undergo technical review only.

Preparatory Access is intended for short-term access to resources, for code-enabling and porting required to prepare proposals for Project Access and to demonstrate the scalability of codes. Applications for Preparatory Access are accepted at any time, with a cut-off date every 3 months. Project Access is intended for individual researchers and research groups including multi-national research groups and can be used for 1-year production runs, as well as for 2-year or 3-year (Multi-Year Access) production runs. Programmatic Access is a pilot in 2014 (10th Call). It is intended to ensure a stable and reliable minimum access to the necessary computational resources for large-scale, long-term projects of very high scientific quality and with a broad European scope, importance and relevance.

PRACE Education & Training

PRACE has an extensive education and training effort for effective use of the RI through PRACE Advanced Training Centers (PATC), seasonal schools, workshops and scientific and industrial seminars throughout Europe. The PATCs update the curriculum yearly following the demand from users and the new technologies and developments. Seasonal Schools target broad HPC audiences, whereas workshops are focused on particular technologies, tools or disciplines or research areas. Education and training material and documents related to the RI are available on the PRACE website.

PRACE and Industry

SHAPE, the SME HPC Adoption Programme in Europe is a pan-European, PRACE-based programme supporting HPC adoption by SMEs. The Programme aims to raise awareness and equip European SMEs with the expertise necessary to take advantage of the innovation possibilities opened up by HPC, thus increasing their competitiveness.

7.2.2 EGI services

EGI provides four solutions to accelerate compute and data intensive research. These solutions are built on top of service, products and intellectual capital that is delivered by the EGI collaboration according to the user requirements.

The EGI Federated Cloud Solution enables to deploy on-demand IT services via standards-based interface onto federate academic and commercial clouds from multiple provider. It can host datasets of public or commercial relevance, and can be flexibly expanded by integrating private cloud providers so to create a hybrid system. RIs can get advanced compute capabilities, virtualised resources to run any environment of choice, cloud storage for easier sharing of data, transparent migration of Virtual Machine images across different providers and a number of support services.

The EGI Federated Cloud offers:

- Single-sign-on authentication and authorization for multiple cloud providers
- Total control over deployed applications
- Elastic resource consumption based on real needs
- Workloads are processed immediately

- The researchers' own cloud infrastructures can be federated in the EGI cloud.

The EGI High Throughput Computing Solution enables to analyse large datasets, or to execute thousands of computational tasks. It is based on a federated infrastructure of hundreds of independent research institutes, universities and organisations delivering computing clusters and high scalability. This solution is aimed at helping individual researchers and research communities that have large-scale distributed data management and computational capacity requirements. With this solution users gain access to their own and externally provided storage, computing and data management services that are made accessible with uniform interfaces. The solution provides a single entry point to this federated pool of services.

The EGI Federated Operations Solution provides technologies, processes and people required to manage the operations of a heterogeneous infrastructure and to integrate resources from multiple independent providers with a lightweight central coordination. Through this solution (or a combination of services of choice) RIs that already own computing and data centres or who will procure these, can integrate the RI resources to the appropriate existing or new Virtual Organisations, or even possibly outsource this effort to their NGIs.

The Federated Operations solution helps RIs that are geographically and/or structurally dispersed and wish to organise themselves for federated service provision.

The Community-Driven Innovation & Support Solution is aimed at helping the individual researchers and the research teams that have problems in accessing and using computational services for their research activity, and is offered to involve RIs in the process of co-designing and evolving the previous three solutions to meet their requirements.

The four solutions are supported by the following two cross-cutting services:

- The Applications database (AppDB): a centralised service that stores information about software tools integrated with the EGI infrastructure, including: ready-to-use scientific applications for many disciplines, science gateways, workflows and software components for developers.
- The Training Marketplace provides a space for trainers and trainees to advertise and look for training events, online courses and training materials on a wide-range of scientific and distributed computing topics. It is a fundamental platform to share and federate knowledge, and foster reuse of training and education material.

The four solutions presented above can be customized according to the user requirements, and the services that together implement the solution can be separately offered to the RIs who have specific interest in one or more of these.

The EGI solutions can be accessed through the following access modes:

1. Policy-based: users are granted access based on policies defined by the EGI resource providers or by EGI.eu; such policies usually apply to resources being offered "free at point of use" to meet some national or EU level objective; for instance, a country may offer free at point of use resources to support national researchers involved in international collaborations.
2. Excellence-driven: users are granted access based on the scientific excellence, originality, quality, technical and ethical feasibility of the work; this access mode is applied by some EGI resource providers; it may be adopted in the future at European level; applications that positively evaluated, receive services free at point of use;

3. Market-driven: users can negotiate a fee to access services either directly with EGI resource providers or indirectly with EGI.eu;
4. Membership-based: RIs can become participants of EGI.eu by paying an annual membership fee (e.g. like CERN or EMBL-EBI); this allows to access a number of benefits including services for federating infrastructures (i.e., the federated operations solution).

The High Throughput Computing solution and Federated Cloud solutions can be accessed either via the following access modes: policy-based, excellence-driven and market-based.

The Federated operations solution can be accessed via the membership-based access mode. The policy-based access mode is possible for non-EU organisations via signing MoUs. The market-based access mode is being developed.

The community driven innovation and support solution can be accessed via all access modes.

7.2.3 HELIX NEBULA services

Since its creation in 2011, Helix Nebula has grown to become a leading public-private partnership between public research actors and cloud service providers. The Initiative has brought together cloud infrastructure providers and large scientific user communities to overcome barriers to adoption of Infrastructure as a Service (IaaS) for scientific use. In addition to offering the production Helix Nebula marketplace platform (HNX[22]), the initiative has undertaken the first joint Pre Commercial Procurement (PCP) tender called Helix Nebula Science Cloud (HNSciCloud). This €5.3 million joint tender, led by CERN, will establish a European hybrid cloud platform that will support high-performance, data-intensive scientific use-cases sponsored by 10 of Europe's leading public research organisations (CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, STFC, SURFSara) and co-funded by the European Commission.

The procured cloud services are being integrated with the procurers' in-house resources and publicly funded e-Infrastructures (GEANT network and EGI Fed Cloud) to provide a hybrid platform for end-users from a wide range of scientific fields[23] including high energy physics, life sciences, astronomy, neutron/photon sciences and the long tail of science. The set of use-cases that will be supported by the procurement include those directly connected to 7 Research Infrastructures that appear in the ESFRI 2016 roadmap, namely:

- Euro-BioImaging: European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences
- CTA: Cherenkov Telescope Array
- BBMRI: Biobanking and BioMolecular resources Research Infrastructure
- ELIXIR: A distributed infrastructure for life-science information
- ESRF Upgrades: Extremely Brilliant Source
- European-XFEL: European X-Ray Free-Electron Laser Facility
- HL-LHC: High-Luminosity Large Hadron Collider

The HNSciCloud tender was published during 2016 and the adjudication process has completed with contracts awarded for the design phase to four consortia:

- T-Systems, Huawei, Cyfronet, Divia

[22] http://hnx.helix-nebula.eu/

[23] http://www.hnscicloud.eu/hnscicloud-user-groups

- IBM
- RHEA Group, T-Systems, exoscale, SixSq
- INDRA Systems, HP, Advania, SixSq

The procured services will be competitively evaluated during prototype and pilot phases with the end-users through 2017 and 2018 after which the most successful services will be commercialised and made generally available.

In parallel, the Helix Nebula initiative is working to build on these results and move "up the stack" towards Platform As A Service (PAAS), Software As A Service (SAAS), Data As A Service (DAAS) activities and prepare the way for Information as a Service (INFOaaS) building the value chains that will create value for both the growing cloud industry, and the scientific user communities, through the reusability and sharing of open data.

7.3 Data Services

7.3.1 EUDAT

EUDAT's vision is data shared and preserved across borders and disciplines and its mission is to enable data stewardship within and between European research communities through the EUDAT Collaborative Data Infrastructure. EUDAT[24] offers a common model and service infrastructure for managing data spanning all European research data centres and community data repositories. EUDAT is the largest pan-European data infrastructure initiative initiated under the EC FP7 programme and is conceived as a network of collaborating, cooperating centres. EUDAT combines the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data and computing centres.

The EUDAT partnership is a network of 35 pan-European trusted organisations – data centres, HPC centres, and research institutions – all closely connected across Europe to offer scientific communities and users reliability, sustainability and confidence in the longevity of research data management services. In September 2016, a 10-year agreement, beyond the EC contract, was signed by 18 of the EUDAT service providers, guaranteeing long term sustainability to communities, infrastructures and their end-users. In addition, EUDAT relies on the best professional IT management practices in use to preserve data (multiple copies, back-up, and recovery plan).

Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT services aim to address the full lifecycle of research data. The current suite of EUDAT data management services includes:

1. A secure and trusted data exchange service[25] (B2DROP) for researchers and scientists to keep their research data synchronized and up-to-date and to exchange with other researchers.
2. A web based service for researchers and communities to store and publish[26] (B2SHARE) small-scale research data coming from diverse contexts.

[24] http://www.eudat.eu

[25] b2drop.eudat.eu/

[26] b2share.eudat.eu/

3. A data management and replication[27] (B2SAFE) service allowing community and departmental repositories to replicate and preserve their research data across EUDAT data nodes.
4. A service to ship large amounts of research data (B2STAGE) between EUDAT data nodes and workspace areas of high-performance computing systems.
5. A metadata catalogue[28] (B2FIND) of research data collections stored in EUDAT data centres and other repositories allowing to find collections of scientific data quickly and easily, irrespective of their origin, discipline or community
6. A single sign-on authorisation and authentication solution[29] (B2ACCESS) for end-users to access EUDAT CDI
7. A service to assign PIDs (B2HANDLE) to different kinds of managed objects stored in the EUDAT CDI

EUDAT is currently organised as a network of centres independent from each other but working within a common framework to develop and operate services on a pan-European level. B2DROP, B2SHARE, and B2FIND are offered free of charge at the point of use to anybody and are available through the web, while B2SAFE and B2STAGE typically require agreements between the user and the service provider. Such agreements can build upon existing relationships between some research communities and data centres, in particular when long standing agreements have been made with national funders to support a specific community through a national centre or a designated service provider. Individual agreements can also be concluded between a research community and individual EUDAT centres offering SLA-based services as part of the EUDAT collaborative framework. Calls for collaboration are regularly held and provide research communities with the possibility to apply for free storage resources during a limited period.

7.3.2 OpenAIRE

OpenAIRE is the European Open Access (OA) e-Infrastructure and enables researchers to deposit research publications and data into Open Access repositories. OpenAIRE also provides support to researchers at the national, institutional and local level to guide them on how to publish in OA and how to manage the long tail of science data within the institution environment. It operates a technical infrastructure that promotes interoperability to data providers through a set of guidelines, and gathers research results deposited in any institutional and thematic literature or data repositories (green OA) or published in OA Journals (gold OA).

Currently the majority of OpenAIRE services are free and do not require registering with a username and password, while in the future other models will be evaluated including subscriptions with added value services and Service Level Agreements.

In particular, the discovery of interlinked services of publications-data-projects-organisations-data providers, along with related provision of reports on project and funder outcomes and on Open Access evaluation (via the dashboard) do not require registration, while the related OpenAIRE APIs (api.openaire.eu) are also free with limited streaming capabilities (<1000 per request). Finally, the linking services, i.e. linking publications-funding-data, require registration.

[27] http://www.eudat.eu/services-support

[28] http://b2find.eudat.eu/

[29] http://www.eudat.eu/services-support/

Zenodo

Zenodo is a research publication and archival service offered as a catch-all for Open Science. Zenodo accepts all research objects, so researchers without easy access to an institutional or a subject repository, can deposit their articles, research data, software and much more. Hosted by CERN, Zenodo exposes its contents through OpenAIRE to the EC reporting channels and offers a range of access policies helping researchers to comply with the Open Access demands from the EC and the European research Councils. Zenodo promotes data sharing through creation of persistent identifiers for articles, research data and software. This service, which offers long-term guarantees, is already used by large projects and RIs to manage part of their data and has been adopted by individual researchers in many fields. It already houses 3,700 datasets and 13,800 software packages, making it the leading issuer of software DOIs in the world.

Zenodo is free for the long tail of Science to store up to 50GB per dataset, with mulitple datasets allowed. Collections of datasets can be built and curated by communities on a self-service basis. Larger files and datasets can be and supported after validation of the use cases, but since infinite space cannot be offereed for free, donations from heavy users are encouraged!

EGI

EGI's solutions also include data infrastructure services, mainly grid storage, cloud storage, file transfer service and file metadata catalogue. Read section 7.2.2 for more details.

Helix Nebula

As a cloud service provider Helix Nebula also offers data services. See section 7.2.3 for more details.

8 Good practices and recommendations

Based on the information in the previous chapters a number of best practices and recommendations can be defined when addressing the e-Infrastructure needs of research infrastructures.

A first general recommendation is that the e-Infrastructure needs and data aspects have to be fully taken into account from the beginning of the research infrastructure study phase and through its whole lifetime.

Recommendations to assess the e-needs and define the ICT environment for the research infrastructure:

- Check existing e-Infrastructures and related services before defining the ICT infrastructure for your research infrastructure.
- Check with existing research infrastructures how they realised their ICT infrastructure;
- Contact existing e-Infrastructures at national and European level and ESFRI Projects and Landmarks - as appropriate - to see if they can accommodate (part of) your e-needs.
- Work to an ICT synergy with other projects to encourage the development of the e-Infrastructure commons. This can include participation in the discussions of the interoperability framework in the research infrastructure disciplinary field and of the generic data sharing one (for instance in the RDA), to make sure that your specific requirements are taken into account.
- Plan your access to ICT resources; e-Infrastructures can be accessed via different access modes that can imply consequences at the budget level.

Recommendations to assess the e-needs instructions about data:

- Take into account the FAIR principles for your research data.
- Define a data management plan that fits your research infrastructure.

9 References

Information about national e-Infrastructures
http://knowledgebase.e-irg.eu/countries
Information GÉANT Association
http://www.geant.org/Pages/Home.aspx
http://www.geant.net/About/partners
GÉANT NREN Compendium
https://compendium.geant.org/
PRACE information
http://www.prace-ri.eu/best-practice-guides/
http://www.prace-ri.eu/user-documentation/
http://www.training.prace-ri.eu/
http://www.prace-ri.eu/hpc-access/shape-programme
EGI information
https://www.egi.eu/services/cloud-compute/
https://www.egi.eu/services/high-throughput-compute/
https://www.egi.eu/internal-services/operations-coordination-and-support/
https://www.egi.eu/internal-services/community-coordination/
https://appdb.egi.eu
https://www.egi.eu/services/

Helix Nebula information
http://www.helix-nebula.eu/
http://hnx.helix-nebula.eu/
Zenodo information
http://www.zenodo.org/
EUDAT information
http://www.eudat.eu/
OPENAIRE information
https://www.openaire.eu/
RDA information
https://europe.rd-alliance.org/
https://europe.rd-alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-yield-knowledge-jobs-and
Data Management policy and plan. Models and guidelines for a data management plan
https://www.icsu-wds.org/services/data-policy
http://www.shef.ac.uk/ris/other/gov-ethics/grippolicy/practices/all/rdmpolicy
http://www.dcc.ac.uk/resources/data-management-plans/checklist
http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdfDPHEP, http://www.dphep.org/
H2020 document Guidelines on Data Management
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

10 Glossary

Term	Explanation
Cloud	Distributed computing and storage infrastructure able to provide services on different levels of abstractions (IaaS, SaaS, PaaS)
CSP	Certification Service Provider
DMP	Data Management Plan
EGI	European Grid Initiative
e-IRG	e-Infrastructure Reflection Group
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT	European Data Infrastructure
GÉANT	Pan-European research and education network that interconnects Europe's National Research and Education Networks (NRENs).
HPC	High Performance Computing
HTC	High Throughput Computing
IP	Internet Protocol
L3VPN	Layer 3 Virtual Private Network
NGI	National Grid Initiative
NREN	National Research and Education Network
OA	Open Access
OpenAIRE	Open Access Infrastructure for Research in Europe
PATC	PRACE Advanced Training Centre
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
RI	Research Infrastructure
SME	Small and Medium Enterprise
TERENA	Trans-European Research and Education Networking Association
Tier-0	Leading-edge supercomputers with very high performance

Tier-1	Supercomputers with high performance; one level less than Tier-0 systems
VO	Virtual Organisation

http://e-irg.eu