e-IRG Guidelines Document

e-IRG e-Infrastructure Reflection Group

Paving the way towards a general purpose European e-Infrastructure

Best Practices for the use of e-Infrastructures by large-scale research infrastructures

Please note this is a historical document.
It has been superseded by:
http://e-IRG.eu/catalogue/eirg-1004
in February 2017

February 24, 2015

http://e-irg.eu

Colophon

ISBN 978-90-823661-0-5

Version 2.0

Table of Contents

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 reinforcement 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, 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, i.e. 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 the e-Infrastructure needs are already considered at the planning stages. Currently, a process is initiated where proposals for new pan-European research infrastructures seek to be included in the updated ESFRI roadmap. Here, information on the needs for e-Infrastructures is requested in the proposal form. The main purpose of this guidelines document is to provide recommendations and best practices concerning e-Infrastructures for large-scale European research infrastructures in their planning and implementation phases. It can also be useful for existing research infrastructures that are planning to update their ICT environment.

The current e-Infrastructure landscape in Europe can be perceived as somewhat scattered, the current 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 available today, and which organisation to contact to get more information.

Sverker Holmgren

e-IRG Chair

February 2015
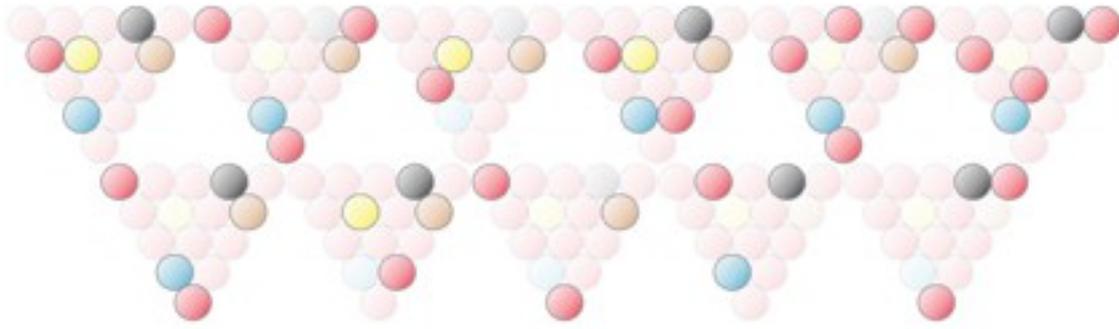
1 Management Summary

This e-IRG guidelines document provides information concerning the use of existing and upcoming e-Infrastructures for the fulfilment of the ICT requirements for large-scale research infrastructures, with a special focus on providing information and recommendations that would be helpful in e.g. the process of writing proposals for the ESFRI 2016 roadmap.

This document contains three main parts.

The first part addresses the e-needs questions that are included in the ESFRI proposal submission form and the corresponding indicators that will be used for the evaluation of the proposal where e-needs stands for the required e-Infrastructure and related services.

The second part discusses a set of recommendations and best practices to be considered when responding to the e-needs questions. In summary the recommendations are:

- Define a data management policy;

- Define a data management plan if appropriate to your RI;

- Ensure that data formats are standardised and contain sufficient information on the data (metadata);

- Define e-Infrastructure solutions consisting of multiple layers, successively adding more specialised higher-level services using standardised interfaces;

- Define and successively move towards a common data storage layer which can effectively serve requirements coming from different RIs. Also here, standardised interfaces and federative approaches should be used to include existing solutions;

- Adopt a global, standardised lowest-level data infrastructure, including e.g. authorisation and authentication and persistent data identifiers; Ensure that quality of the e-Infrastructure services and the data security is delivered at a level which is relevant for the data at hand;

- Pay attention to the sustainability of your data, also after the end of the project.

The best practices can be summarised as follows:

- Check existing e-Infrastructures and related services for their use before defining the ICT infrastructure for your Research Infrastructure;

- Check with existing RIs how they realised their ICT infrastructure;

- Contact existing e-Infrastructures and ESFRI projects at national level and/or European level as appropriate;

- Work to an ICT synergy with other projects to encourage the development of the e-Infrastructure commons;

- Pay attention to the interoperability of services and data;

- Plan your access to ICT resources.

The third part provides background information to be used in the process of answering the questions and to follow up the recommendations and best practices presented in this document. We describe mainly the e-Infrastructures at European level but are aware that e-Infrastructures at national or even 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.
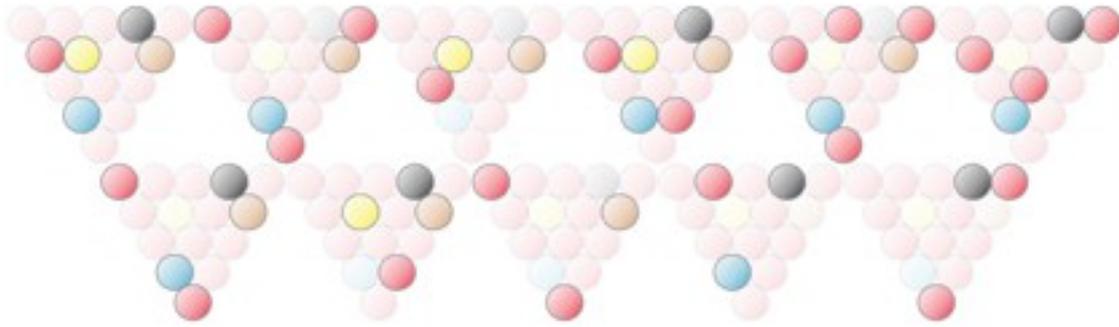
2 Introduction

The aim of this guidelines document is to guide large-scale research infrastructures in the process of planning and of implementing in a later phase 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). Large-scale international research infrastructures normally have significant networking, computing and data management needs. In the past such research infrastructures often had to establish their "own" e-Infrastructure components and implement and operate 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 to use the 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 very 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 many of the stakeholders to grasp the needs of users and the advantages of using existing e-Infrastructures and related services.

The ESFRI call for proposals for the 2016 roadmap has taken up this issue. Section 4 "e-needs", in "Part B: scientific impact, pan European relevance, socioeconomic impact & e-needs" of the online submission form for an ESFRI proposal, lists questions relevant to the specification of the e-Infrastructure requirements of the proposed ESFRI research infrastructure. It is important for all ESFRI project proposers to reflect on their needs for networking, computing, data and related services and to introduce their plans on these matters in their proposal. Indicators have been defined that will be used to evaluate the answers on the e-needs questions.

To be able to present the answers to the e-needs questions it is important for the proposers of ESFRI initiatives to have an understanding of the e-Infrastructure landscape in Europe and in the individual countries. Hence we provide a short overview of the basic e-Infrastructure components focussed on the European scale. However, note that in many cases also e-Infrastructures at a national level can be of great importance to ESFRI initiatives and that often a coordination exists between the

national and European level. **It is not possible to elaborate in all details in this document on these national e-Infrastructures but instead pointers to them are provided in the form of references in the e-IRG Knowledge Base.** In general, contact information and pointers to more information about e-Infrastructures are presented in the References section.

Even as this guidelines document is in the first place aimed at research infrastructures in their planning phase it may also be useful for existing research infrastructures that plan to update their e-Infrastructure and/or extend their e-Infrastructure services.
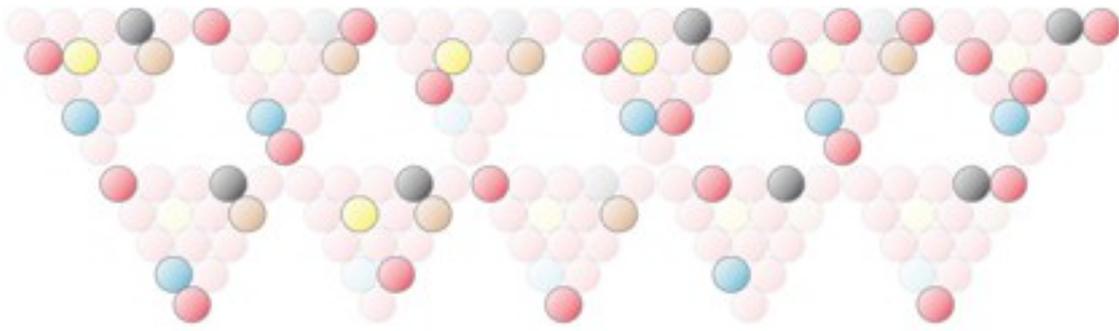
3 Questions and indicators concerning the description of e-Infrastructure services for the ex-ante evaluation of a research infrastructure

A number of questions concerning e-Infrastructure services have been drafted and included in the ESFRI call for proposals (see the online submission form, referenced below). These questions are listed below followed by a list of the related indicators with a corresponding explanation. These indicators will be used in the evaluation of the ESFRI proposals.

3.1 Questions

The following questions are part of the ESFRI *Online Submission Form: Proposals for the 2016 ESFRI Roadmap* [1] and aim at reflecting the e-needs of the ESFRI proposal.

Nr.	Question
4.1	What will be the data management and open data policy of the RI? (e.g. would data become accessible from a repository to the public? Would the RI be interfaced to e-Infrastructures for science?)
4.2	What is the plan for supporting advanced data management and how will it be funded?
4.3	What is needed (if applicable) from external e-Infrastructure services (resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc.)?
4.4	Will the RI contribute to the development of e-Infrastructure commons in the field or in general?
4.5	Will the RI policy on data include training services for "data practitioners" to enable the effective use of data repositories and data analysis tools by non-scientists?

---

[1] http://ec.europa.eu/research/infrastructures/pdf/20141024-115124_ESFRI%20Online_Submission_Form_21102014.doc.pdf

3.2   Indicators linked to the questions

Section 4. e-NEEDS			
Objective	Indicator High, Medium, Low	Comment on interpretation of the indicator	Date source
4.1	Existence of data policy.	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. Answer might refer to existing and documented policies.	Project management
4.2	Existence of data management plan. Existence of funding scheme.	Procedures and resources needed for data collection, documentation, storage and access. Procedures and resources needed for long-term preservation, including usability, availability, time-frame, etc. Plan for funding.	Project management
4.3	List of e-Infrastructure resources. Explanation of their provision.	List of e-Infrastructure resources (local, national, pan-European, global) needed by the RI to fulfil its purpose but that are not provided by the RI, and explanation how these will be made available.	Project management
4.4	Assessment of RI relation to e-Infrastructure Commons.	Relation to components of e-Infrastructure commons e.g. usage of and contribution to interoperability frameworks/standards and interoperable services; see Paragraph 3.1 and the e-IRG White Paper 2013[2]	Project management
4.5	Training plan and target audiences.	"Non-scientists" means researchers in other domains than usually catered for by the RI, and "data practitioners" means a person who can use existing services for research and innovation without expert knowledge of the RI itself.	Project management

---

[2] http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf
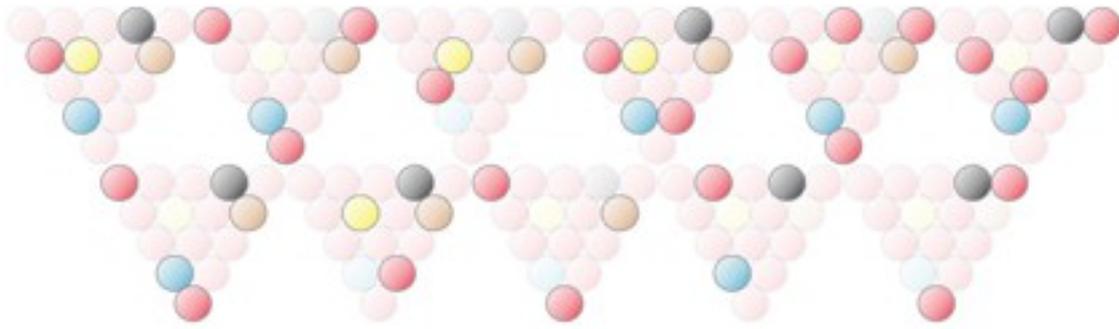
4 Best practices and recommendations

Based on the discussions and summaries in section 5, a number of best practices and recommendations can be defined when analysing 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. The needs and the required resources have to be assessed and the data policy, including the data sharing rules, and the data life cycle, have to be defined.

As a fundament for research infrastructures, sustainable e-Infrastructure services for enabling access to storing, preserving and curating large amounts of data need to be in place. To ensure that data will be available across borders and disciplinary domains, research infrastructures and e-Infrastructure providers are recommended to take appropriate steps to:

- Ensure that data formats are standardised and contain sufficient information on the data (metadata) to enable global usage within the discipline, across disciplines, and in new research settings that could possibly not be envisaged at the time of creation of the data;

- Build e-Infrastructure solutions consisting of multiple layers, successively adding more specialised higher-level services using standardised interfaces. Here, different layers can be provided by different actors (e.g. EUDAT, OpenAIRE, national projects);

- Define and successively move towards a common data storage layer, which can effectively serve requirements coming from different RIs. Also here, standardised interfaces and federative approaches should be used to include existing solutions;

- Adopt a global, standardised lowest-level data infrastructure, including e.g. authorisation and authentication and persistent data identifiers. Here, federative approaches could be used to include existing solutions, e.g. eduGAIN;

- Ensure that quality of the e-Infrastructure services and the data security is delivered at a level, which is relevant for the data at hand.

In addition to the above data related recommendations the following best practices should be taken into account:

- Check existing e-Infrastructures and related services before defining the ICT infrastructure for your research infrastructure;

- Check with existing RIs how they realized their ICT infrastructure;

- Contact existing e-Infrastructures and ESFRI projects at national level and/or European level as appropriate;

- 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 the generic one, to make sure that your specific requirements are taken into account;

- Pay attention to the interoperability of services and data;

- Build a data management plan. A general management plan may not be readily available however several examples of data management plans are available, see the References chapter for more information;

- Plan your access to ICT resources; e-Infrastructures can be accessed via different access modes that can imply consequences at the budget level;

- Pay attention to the sustainability of your data, also after the end of the project.
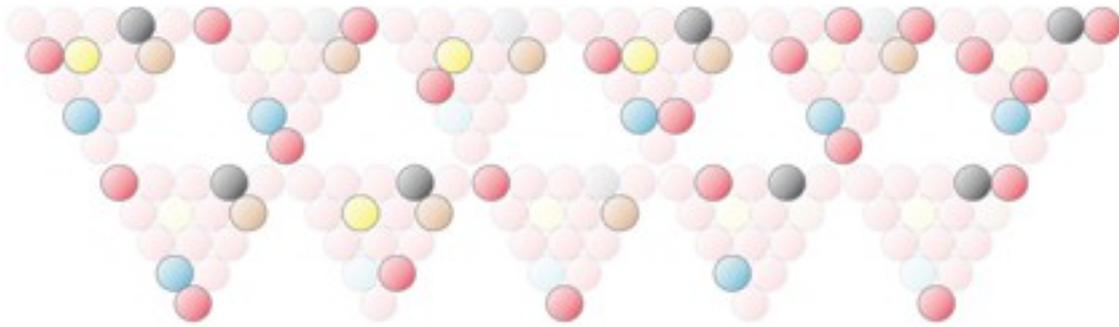
5 The general e-Infrastructures landscape

The current landscape of e-Infrastructures is composed of basic components, namely network, computing and data infrastructures, along with a series of other services including instrumentation, software and middleware, such as authentication and authorisation infrastructures, and other supporting and collaboration tools and resources. The EC made a comprehensive background analysis of e-Infrastructures in 2009: the "Communication ICT Infrastructures for e-Science"[3]. While this document is now outdated for what concerns the existing "production" e-Infrastructures it still gives a good idea of the importance of e-Infrastructures. The past years showed that networking and computing infrastructures are rather well-established and provide full production services while the data infrastructures are still at project level and start to offer services.

e-Infrastructures at national and European level have almost developed jointly and are strongly linked together. However, each national e-Infrastructure has, in general, its own rules, policies and access procedures. The corresponding European e-Infrastructure landscape may thus suffer from a lack of uniformity due to these different national regulations.

It is also worthwhile to mention cloud services. Cloud services means services made available to users on demand via the network from a cloud service provider's servers. Cloud services are designed to provide easy, scalable access to applications, resources and services, and are fully managed by a cloud services provider.

A cloud service can dynamically scale to meet the needs of its users, and because the service provider supplies the hardware and software necessary for the service, there's no need for a company to provision or deploy its own resources or allocate IT staff to manage the service. Cloud services can include computing, data and software services. Different cloud services are taken up in the computing and data services of chapter 6.

Below we briefly introduce major e-Infrastructures, e-Infrastructure projects and related services at European level. The national e-Infrastructures are certainly as important as their international equivalents. However it is impossible to describe in this document all national e-Infrastructures and related services, therefore we point to the **country information** that is stored in the **e-IRG Knowledge Base**[4] and to the information about national initiatives that is mentioned in the documentation of the European e-Infrastructures.

---

[3]COM/2009/0108 final, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - ICT Infrastructures for e-Science
[4]http://knowledgebase.e-irg.eu/countries

5.1   Networking infrastructure

GÉANT, the pan-European Research and Education Network, provides a high performance network infrastructure supporting connections across 43 European countries, from Iceland in the North to Israel in the South. In addition the GÉANT network provides international connections to partner networks worldwide enabling international Research and Education collaboration. GÉANT provides interconnectivity between Europe's National Research and Education Networks (NRENs).

Illustration 1: GÉANT European connections
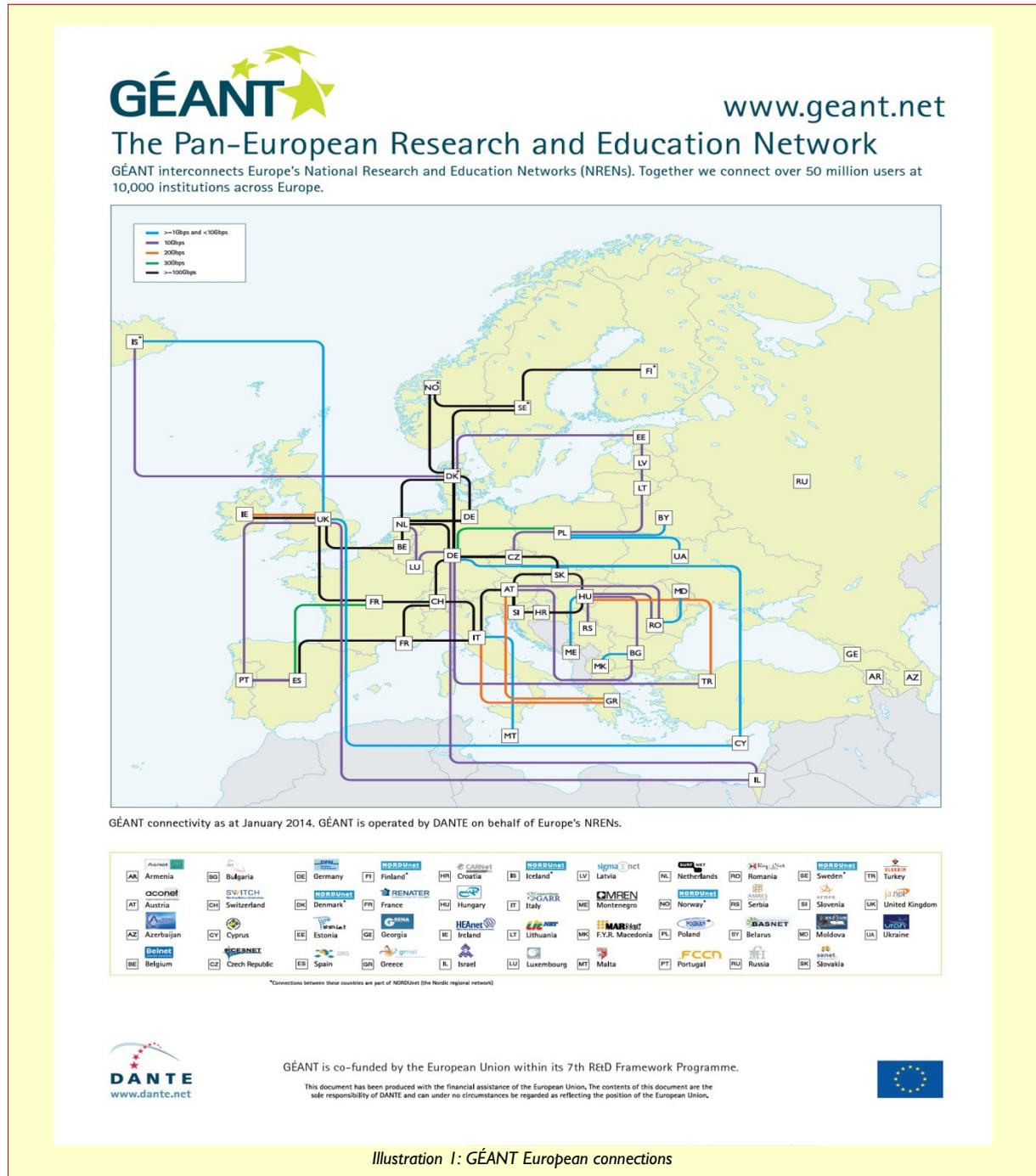
The NRENs connect universities, research institutes, and sometimes other public institutions in their country. Most large-scale research infrastructures can connect to the local NREN and thus access GÉANT enabling world-wide communications. Projects can also work with their NREN and GÉANT for international 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.

5.2   Computing infrastructures

The most well-established pan-European computing infrastructures are the European high throughput computing and federated cloud infrastructures within the European Grid Infrastructure (EGI) and the HPC infrastructure represented by Partnership for Advanced Computing in Europe - (PRACE), respectively operated by the organisations EGI.eu and PRACE AISBL. While EGI focuses on high throughput computing and cloud computing, PRACE offers access to world-class high performance capability computing facilities and services.

EGI and PRACE offer complementary access modes. PRACE resources are allocated twice a year based on a unified peer-review process, while EGI provides solutions both for the federation of the RI data and compute services, and the access to externally provided resources according to various access models. Both have very often national partners in the form of NGIs (national grid initiatives) for EGI and in governmental representative organisations for PRACE. In some countries the national partners are the same for EGI and PRACE, in other countries they are not.

Both e-Infrastructures are ready to talk with consortia that prepare a large-scale research infrastructure to understand needs and find out how this matches with their policies or if it may require policy change. Besides EGI and PRACE, the Helix Nebula Partnership has created a federated cloud services market place and GÉANT also provides cloud services via the NRENs. We can observe also many cloud related national initiatives which aim at supporting large European research infrastructure on a country- or region-wide scale.

5.3   Data infrastructures

Data infrastructures are not yet as well-established as the basic networking and computing infrastructures in the European scene. However, significant steps have been made in the areas of lower-level data services (such as storage and replication) through the EUDAT project and access to publications and other research outputs through the OpenAIRE projects. In addition, much effort goes to the definition and development of common or interoperable data formats, metadata and data management services to enable data interoperability and sharing.

The current undertakings aim at the realisation of an ecosystem with the appropriate technical and social channels for openly sharing of data at a multidisciplinary and global level. The latter activities are part of the Research Data Alliance (RDA) initiative, a bottom-up organisation with constituents in different regions (such as RDA Europe) and countries, composed of Working and Interest Groups. The goal of RDA is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange. This is achieved through the development, adoption, and deployment of infrastructure, policy, practice, standards, and other deliverables. The emphasis of RDA is adoption and deployment as the result of focussed, short-term effort.

5.4   Middleware, tools and related infrastructures

The basic e-Infrastructure services need to be expanded with other services and can include instrumentation, software and middleware services. Examples are collaboration tools and the authentication and authorisation infrastructures.

In the highly distributed and mobile world of high capacity and ubiquitous network access, researchers freely collaborate across boundaries. In this environment, research communities need to

manage access to their services from participants in many organisations and individual researchers need to easily and securely access multiple tools, services and datasets. Trust and identity therefore take up a pivotal position as dedicated layer in the e-Infrastructure eco-system.

Federated Authorisation and Authentication Infrastructures simplify access to inter-organisational resources required by the modern researcher, allowing controlled and secure access to diversely located resources. This provides researchers with a trustworthy environment in which to preserve integrity of their work without compromising their flexibility.

Well established Federated Authorisation and Authentication Infrastructures (national identity federations and eduGAIN), access services infrastructures (eduroam), digital certificates (PKI), security (CERT) and other related services and future developments seamlessly enable safe and secure research throughout Europe and beyond.

These trust and identity infrastructures form a layer connecting the power of the network with computing infrastructure and the value of data infrastructure. The Middleware layer is an essential building block to achieve the vision of a truly open, trustworthy and secure European Research Area (ERA).

GÉANT provides the user access services eduGAIN and eduPKI and application services including eduroam® and eduCONF. Note that the national parts of these services are very important.
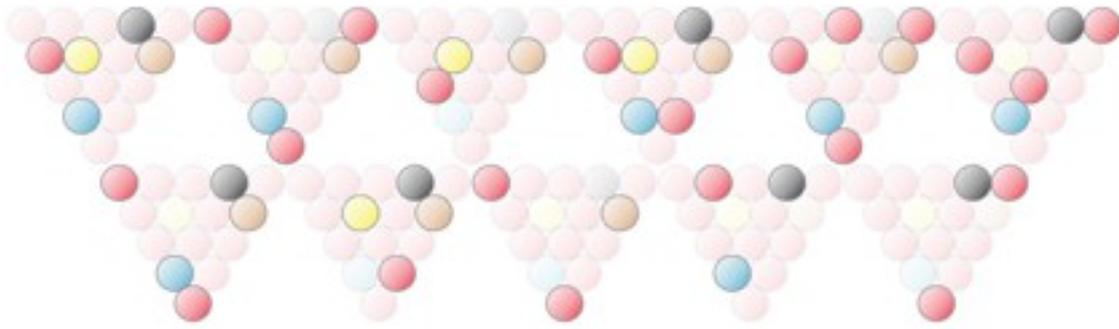
6 An e-Infrastructure Commons

6.1 Definition

e-IRG uses the metaphor of the Commons for the e-Infrastructure resources, which among others refer to networking, computing, storage, data and software, along with digital tools and collaboration opportunities. In the e-IRG Roadmap 2013, e-IRG outlines Europe's need for a single "e-Infrastructure Commons" for knowledge, innovation and science. The e-Infrastructure Commons is the political, technological, and administrative framework for an easy and cost-effective shared use of distributed electronic resources across Europe.

Illustration 2: The e-Infrastructure Commons provides a common interface for research infrastructures to access e-Infrastructures.
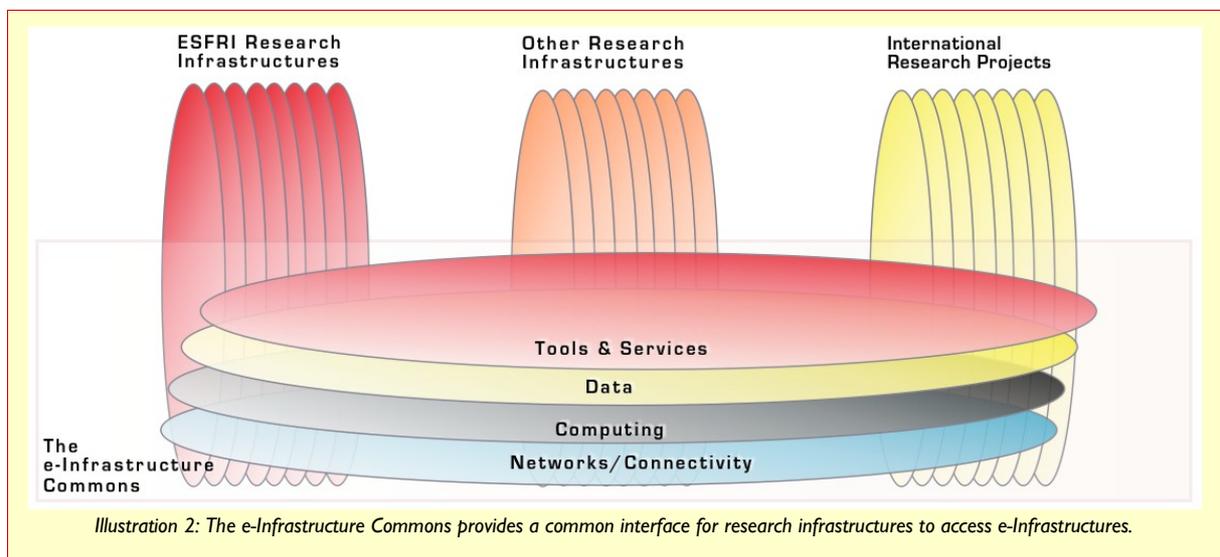
The e-Infrastructure Commons can be thought of as a living ecosystem that is open and accessible to European researchers and scientists, and continuously adapts to the changing requirements of research. Breaking through the limitations of today's separate e-Infrastructure components, services and governance, an "e-Infrastructure Commons" should liberate scientists from the often complex and distracting business of "computing". They need services that are coherent, managed and above all integrated or initially interoperable, so that they can get on with the business of science. However, it

is not the idea that there will be just "one way" of supplying or using any service through an "efficient" mandated or "voluntary" monopoly. It is important that e-Infrastructure providers are encouraged to be open and competitive and above all not "institutionalised". Allowing e-Infrastructures to evolve is important: keywords here are open competition, collaboration but also technological innovation.

Such an e-Infrastructure Commons can only be established through a joint and truly common strategic effort between user communities and the organisations involved in the development and operation of e-Infrastructures, both at the national and at the European level. It is primordial that large-scale research infrastructures contribute to the realisation of this e-Infrastructure commons to their own benefit and that of science in general. They can do so by carefully considering to use existing e-Infrastructures services to fulfil their ICT requirements and by collaborating with other research projects and research communities.

6.2 A range of e-Infrastructure services for large-scale research infrastructures

e-Infrastructure services and access to e-Infrastructures are described in detail in the different 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

6.2.1 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 network services offered by GÉANT include Layer 2 and Layer 3 (IP) services at speeds of up to 100 Gbps. The portfolio includes;

GÉANT Service	Description
GÉANT IP	The network's standard connectivity service providing robust, high-bandwidth access to and across the shared European Internet Protocol (IP) infrastructure.
GÉANT L3VPN	GÉANT L3VPN offers the reliability and flexibility of the award winning GÉANT IP service combined with the additional privacy of a Virtual Private Network (VPN).
GÉANT Plus	Provides virtualised point-to-point connectivity across the GÉANT high-speed backbone to support the most demanding of international data requirements.
GÉANT Lambda	A bespoke service designed to cater for longer term, extremely data-intensive network connectivity requirements, providing a dedicated services up to the full 100Gbps bandwidth.
GÉANT Open	GÉANT Open has been designed as a service to allow NRENs and approved commercial organisations to exchange connectivity in a highly efficient and flexible manner.
GÉANT Bandwidth on Demand	A co-provisioned global multi-domain service that can automatically establish a circuit in minutes, delivering the bandwidth users need, when they need it.

In addition, GÉANT manages peering arrangements with a wide range of other networks and providers to increase the efficiency of IP interconnections. Together, this portfolio offers the R&E community a comprehensive range of networking solutions to support even the most demanding requirements. Descriptions of these services may be found on the GÉANT website at http://www.geant.net/services.

Management and Support

The connectivity delivered by GÉANT is supported by a comprehensive range of network monitoring and management services. These 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.

GÉANT Support Service	Description
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. • Securing the GÉANT Service Area network elements through design and implementation of recommended access and usage policies. • Building proactive security services based on incident databases, anomaly detection tools and common procedures for mitigation of denial-of-service attacks. • Defining a common approach and processes for coordinating responses to security issues.

•

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.

Cloud services

GÉANT is actively helping NRENs to deliver cloud services to their communities, with the right conditions of use. 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.

The goal is to acquire and manage the delivery of services from providers to the pan-European GÉANT community. These parties can be commercial vendors as well as NRENs, and other research and education communities. The goal is an attractive, well-balanced cloud portfolio of cloud services published in online GÉANT Cloud Catalogue, an informative list of cloud providers.

GÉANT is currently working with CSPs to ensure that cloud services meet the needs of the users, Campuses and NRENs.

GÉANT is a collaboration of Europe's NRENs who together connect almost every university and research institute in Europe and through GÉANT's global partnerships, to most other world regions too.

At a user level, services developed and operated by GÉANT are accessed through the local NREN. NRENs provide services to the not-for-profit research and education sector, although some NRENs also connect government departments, industry research centres and other organisations too.

Depending on national funding models, the NREN charging policies vary on a country-by-country basis, but the majority offer access to the GÉANT backbone free-at-the-point-of-use to those institutions connected to the NREN.

6.2.2   Computing services

As mentioned in 2.2 several computing services are available at international and national levels. Below services at international level are briefly described and links with corresponding national services are mentioned. More references are given in Annex 1.

PRACE Services

PRACE provides HPC time to researchers in Europe on a set of 6 Tier-0 and Tier-1 systems of its partners provided by 4 Hosting Members (France, Germany, Italy and Spain). In addition PRACE also offers high value services like an extended training program through 6 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.

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 authorisation 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:

EGI Access mode	Description
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
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
Market-driven:	users can negotiate a fee to access services either directly with EGI resource providers or indirectly with EGI.eu
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, excellent-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.

HELIX NEBULA services

The Helix Nebula initiative 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, including computing services, through the innovative open source broker technology. Federation of public sector data centre resources from within the European Grid Infrastructure (EGI) with HNX may be possible based on users' demand. Access to all service providers is possible via the GÉANT network.

HNX is open to new cloud providers that are able to participate competitively in-line with European regulations and with a suitable quality of service. More commercial cloud providers from various EU member states have recently joined the Helix Nebula Initiative and declared their intent to offer services.

The Cloud services are offered to the global research community, for both publicly-funded and commercial organizations of diverse sectors. Services from the commercial service providers are available of a pay-per-use basis (either on-demand or reserved in advance). Each cloud service provider sets their own prices (e.g. for cloudsigma: https://zrh.cloudsigma.com/ui/#/marketplace).

6.2.3 Data Services

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

The RIs, such as the initiatives 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 RIs 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.

Data infrastructures built by RIs 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.

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 (typically instruments / laboratories) where the users (research projects) generate the data need to clarify which entities should actually develop the data management plan. 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 RI.

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.

Data infrastructures and related services

EUDAT

EUDAT is the largest pan-European data infrastructure initiative initiated under the EC FP7 programme and is set to move towards a sustainable research data infrastructure. 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 B2 services is:

EUDAT Service	Description
B2DROP	a secure and trusted data exchange service for researchers and scientists to keep their research data synchronized and up-to-date and to exchange with other researchers
B2SHARE	a web based service for researchers and communities to store and share small-scale research data coming from diverse contexts
B2SAFE	a data management and replication service allowing community and departmental repositories to replicate and preserve their research data across EUDAT data nodes
B2STAGE	a service to ship large amounts of research data between EUDAT data nodes and workspace areas of high-performance computing systems
B2FIND	a metadata catalogue 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

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.

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 (http://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.

Zenodo

If researchers have no access to an institutional or a subject repository, Zenodo will enable them to deposit their articles, research data and software. Zenodo, hosted by CERN, exposes its contents to OpenAIRE 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 has been extended with important features that improve data sharing, such as the 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.

Zenodo is free for the long tail of Science. In order to offer services to the more resource hungry research, a ceiling to the free slice and offer paid for slices above will be introduced, according to the business model developed within the sustainability plan. A cost level for storage above the ceiling is not yet established but that will follow a non-profit cost recovery approach and be introduced during 2015.

EGI

EGI's solutions also include data infrastructure services, mainly grid storage, cloud storage, file transfer service and file metadata catalogue. Read section 6.2.2.2 for more details.

Helix Nebula

As a cloud service provider Helix Nebula also offers data services. See section 6.2.2.3 for more details.

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 RI. 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*[5]. 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.

RDA published very recently its report *The Data Harvest: How sharing research data can yield knowledge, jobs and growth*[6] in which they call for the sharing of data over vast distances, across institutions and disciplines.

The e-IRG issued already a set of recommendations about data management in its *Blue Paper on Data Management*[7] and *White Paper 2013*[8].

### 6.2.4 Middleware, tools and related infrastructures

As researchers operate in an increasingly mobile, distributed way, it is important that the integrity of their work is protected without compromising their flexibility. Trust infrastructures form a layer connecting the power of the network infrastructure with the value of data infrastructure. A number of user access and application services for the research community are provided by GÉANT, including eduroam, eduGAIN, eduCONF and eduPKI.

edu… Service	Description
eduGAIN	interconnects identity federations around the world, simplifying access to content, services and resources for the global research and education community. It enables the trustworthy exchange of information related to identity, authentication and authorisation across national boundaries. Research communities are empowered to control who accesses their services, and campuses to manage identities, maximising trust while minimising user management. http://edugain.geant.net
eduroam	enables researchers to work flexibly from home or visited institutions in over 70 countries worldwide. This enables flexible, secure, collaboration while minimising the operational management for campuses and research institutes and removing the need for users to maintain multiple connection profiles when they travel. Eduroam supports nearly 5,000,000 users a day with over 500,000 international authentications daily. http://eduroam.org
eduCONF	supports a high quality videoconferencing and collaboration environment that will be easily accessible to research and academic end-users. Effective Video Conferencing can save time, money and improve teamwork. With eduCONF users will be able to find and use videoconferencing facilities across Europe. eduCONF services are helping to provide the tools for an effective pan-European videoconference infrastructure, including tools for locating, validating and

[5]http://zenodo.org/record/7636
[6]https://europe.rd-alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-yield-knowledge-jobs-and
[7]http://e-irg.eu/documents/10920/238805/e-irg-blue_paper_on_data_management_v_final.pdf
[8]http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf

	certifying rooms, managing and monitoring a videoconferencing network http://educonf.geant.net
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
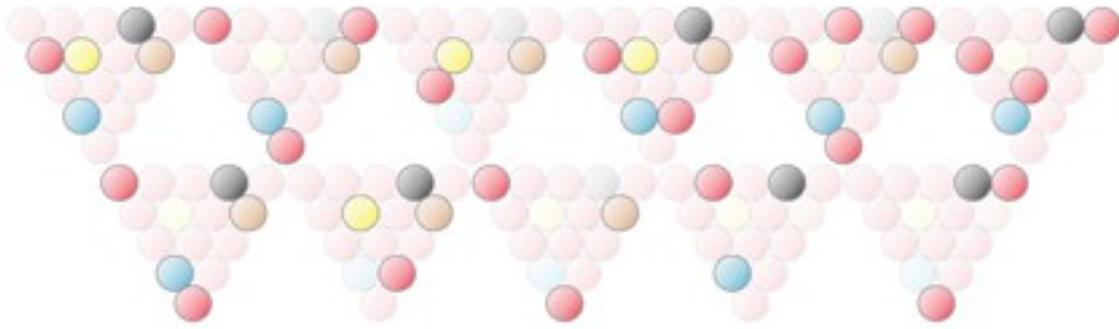
7 Use Cases

Of course, many European research infrastructures and research projects already today are successfully using e-Infrastructures. However, for some e-Infrastructures it might be evident that they are used by universities and research organisations while it is not always clear if use by research infrastructures is also possible. Below we point to success stories in the use of e-Infrastructures by research projects.

Use case	Reference
PRACE use	http://www.prace-ri.eu/prace-the-scientific-case-for-hpc/
EGI use	http://www.egi.eu/export/sites/egi/news-and-media/publications/EGI_Case_studies.pdf
EUDAT use	http://www.eudat.eu/news-media/published-articles/data-services-technology-expertise-community-perspective http://www.eudat.eu/news-media/published-articles/going-dutch-your-data http://www.eudat.eu/news-media/published-articles/interview-massimo-cocco-epos
GÉANT use	http://www.geant.net/Users/Pages/home.aspx
HELIX NEBULA use	http://www.helix-nebula.eu/usecases http://www.helix-nebula.eu/publications/deliverables/d92-strategic-plan-scientific-cloud-computing-infrastructure-europe-three
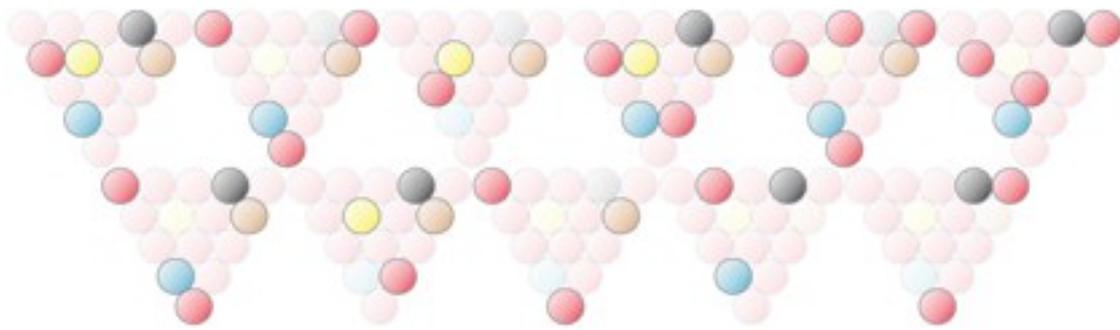
8 References

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
TERENA NREN Compendium
https://www.terena.org/activities/compendium/index.php?showyear=2014
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
http://www.egi.eu/solutions/fed-cloud/
http://www.egi.eu/solutions/htc/
http://www.egi.eu/solutions/fed-ops/
http://www.egi.eu/solutions/community-innovation-support/
https://appdb.egi.eu
http://www.egi.eu/services/training_marketplace/index.html

| |
|---|
| http://operations-portal.egi.eu/http://www.egi.eu/solutions |
| **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 |
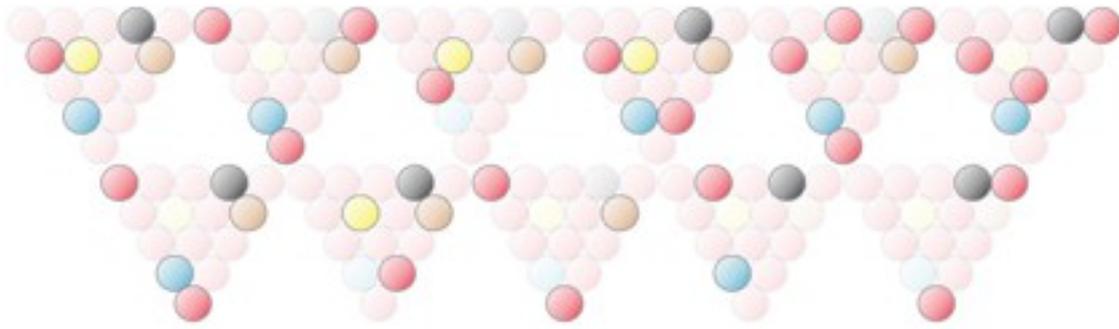
9 Glossary

Term	Explanation
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