



# **e-IRG White Paper 2014**

Early draft version  
26 January 2015

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

Foreword .....	1
1. Management Summary .....	1
2. Introduction .....	2
3. Indicators concerning the description of e-Infrastructure services for the ex-ante evaluation of a research infrastructure.....	3
3.1 Questions .....	3
3.2 Indicators linked to the questions.....	4
4 The general e-Infrastructures landscape.....	5
4.1 Networking infrastructure.....	6
4.2 Computing infrastructures.....	7
4.3 Data infrastructures.....	7
5 An e-Infrastructure commons.....	8
5.1 Definition .....	8
5.2 A range of e-Infrastructure services for large-scale research infrastructures .....	9
5.2.1 Networking services.....	9
5.2.2 Computing services .....	9
5.2.3 Data Services .....	12
6 Best practices and recommendations .....	15
7 Use Cases .....	16
8 References.....	17
9 Glossary.....	18

Draft

# Foreword

## 1. Management Summary

This White Paper contains three main parts to help define the ESFRI project proposer e-Needs.

The first part is formed by 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. Those indicators are:

- Existence of data policy;
- Existence of data management plan, Existence of funding scheme;
- List of e-Infrastructure resources, Explanation of their provision;
- Assessment of RI relation to e-Infrastructure Commons;
- Training plan and target audiences.

To reply to the e-Needs questions a set of recommendations and best practises have been defined. The first recommendations pertain to the data issues in the project:

- Build a data management policy.:
- Build a data management plan if appropriate to your RI;
- Ensure that data formats are standardised and contain sufficient information on the data (metadata);
- Build 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.

Common best practises can be summarised as follows:

- 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;
- Pay attention to the interoperability of services and data;
- Plan your access to ICT resources.

The third part gives all background information needed to answer the questions and to follow the recommendations. As repeated several time throughout the document, we described 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, e-IRG can also be asked to help you with contacting e-Infrastructures.

## **2. Introduction**

The aim of this White Paper is to guide large-scale research infrastructures to the use of national and European electronic Infrastructures (e-Infrastructures). Large-scale research infrastructures more than often have large-scale networking, computing and data management needs and in the past large-scale research infrastructures often established their “own” e-infrastructures and related services. This was done to provide the best possible services adapted to the particular research infrastructure environment and users’ requirements. In the meantime national and European e-Infrastructures have evolved and offer a broad range of services, taking into consideration to some extent user communities’ needs and requirements.

Much effort has been done to bring research infrastructures and e-infrastructures together for their mutual benefit, but it might still be difficult for most of the stakeholders to grasp the advantages of using existing e-infrastructures and related services.

The ESFRI call for proposals 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 the questions relevant to the specification of the e-infrastructure requirements of the project. 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 project proposers to have an understanding of the e-infrastructure landscape in Europe and in the individual countries. Hence we provide you with a short overview of the basic e-Infrastructure components mainly on a European scale. Bear in mind that e-Infrastructures at national level are of equal importance. However it is not possible to elaborate here on these national e-Infrastructures but pointers to them are provided. In General, contact information and pointers to more information about each e-Infrastructure is presented in the References section.

Finally the best practises and recommendations for the proposers of ESFRI projects are proposed in Section 6.

### **3. 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). Those 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 proposal.

#### **3.1 Questions**

The following questions are part of the ESFRI *Online Submission Form: Proposals for the 2016 ESFRI Roadmap*<sup>1</sup> and reflect the e-Needs of applicants.

- 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?

---

<sup>1</sup> [http://ec.europa.eu/research/infrastructures/pdf/20141024-115124 ESFRI%20Online Submission Form 21102014.doc.pdf#view=fit&pagemode=none](http://ec.europa.eu/research/infrastructures/pdf/20141024-115124_ESFRI%20Online_Submission_Form_21102014.doc.pdf#view=fit&pagemode=none)

### 3.2 Indicators linked to the questions

Objective	Indicator High, Medium, Low	Comment on interpretation of the indicator	Date source
<b>4. e-needs</b>			
4.1	<b>Existence of data policy.</b>	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	<b>Existence of data management plan.</b> <b>Existence of funding scheme.</b>	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	<b>List of e-Infrastructure resources.</b> <b>Explanation of their provision.</b>	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	<b>Assessment of RI relation to e-Infrastructure Commons.</b>	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 <a href="#">e-IRG White Paper 2013</a> <sup>2</sup>	Project management
4.5	<b>Training plan and target audiences.</b>	“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

<sup>2</sup> <http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf/ce8a2253-aebd-4cbe-9a93-4709a1166214>

## 4 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”<sup>3</sup>. 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 but that the data ones 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 its own rules, policies and access procedures, although same basic principles are applied overall. The corresponding European e-Infrastructure landscape may thus suffer from a lack of uniformity due to these different national regulations.

Below we briefly introduce major e-Infrastructures and e-Infrastructure projects 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<sup>4</sup> and to the information about national initiatives that is mentioned in the documentation of the European e-Infrastructures.

---

<sup>3</sup> [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](#)

<sup>4</sup> <http://knowledgebase.e-irg.eu/countries>

## 4.1 Networking infrastructure

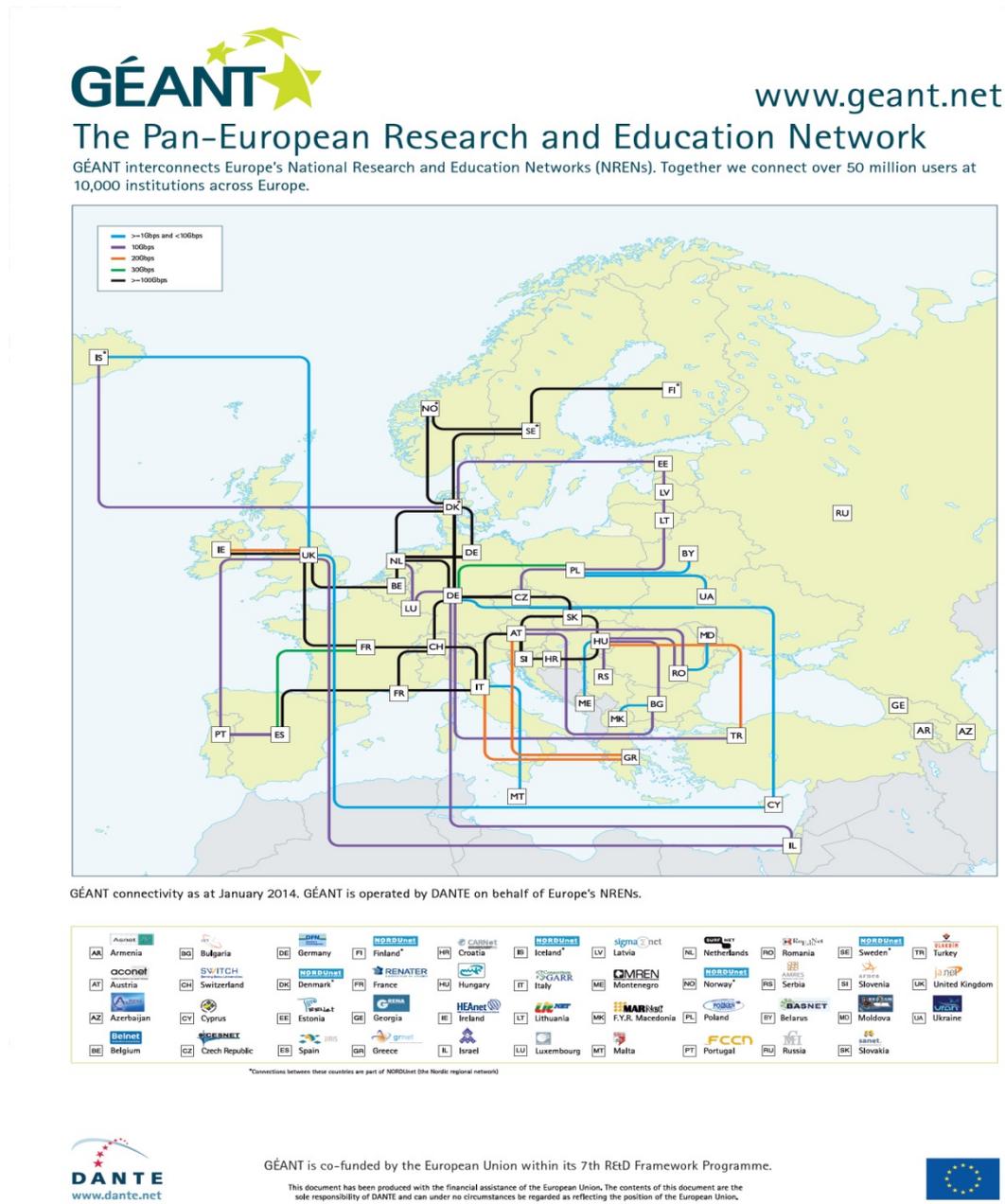


Figure 1 GÉANT European connections

GÉANT, the pan-European Research and Education Network, as pictured above, ensures stable network connections across 43 European countries, from Iceland in the North to Israel in the South. And the networking does not stop at the European borders – international connections are in place connecting GÉANT to 65 countries in other continents. In fact GÉANT interconnects Europe’s National Research and Education Networks (NRENs). The NRENs connect universities, research institutes, and sometimes other public institutions in their country. The connection to a NREN is a local policy matter and can differ from country to country. Most large-scale research infrastructures can connect to the local NREN and thus access GÉANT enabling world-wide

communications. Projects can also turn to GÉANT for international point to point links to connect parts of the research infrastructure that are distributed over Europe or beyond. Organisations that want to make use of the NRENs or GÉANT should contact their local NREN to understand the services offered, the terms and conditions. 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.

## **4.2 Computing infrastructures**

The well-known pan-European computing infrastructures are the European High Throughput Computing and Federated Cloud infrastructures (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 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. 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 national initiatives cloud related which can support large European initiatives of RI on a micro scale (country).

## **4.3 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 An e-Infrastructure commons

### 5.1 Definition

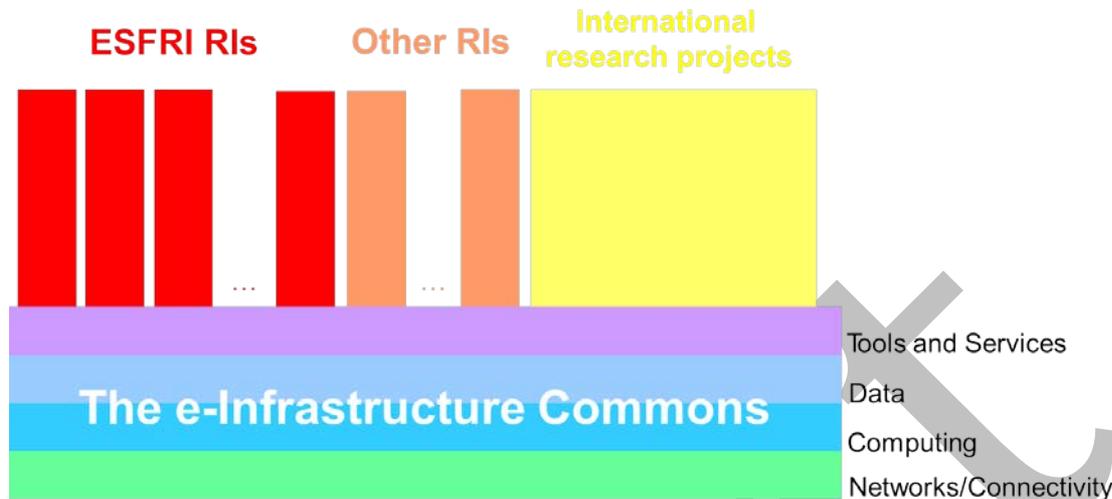


Figure 2 the e-Infrastructure Commons

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. 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.

## **5.2 A range of e-Infrastructure services for large-scale research infrastructures**

### **5.2.1 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 TERENA 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 GÉANT IP, GÉANT L3VPN, GÉANT Plus, GÉANT Lambda, GÉANT Open and GÉANT Bandwidth on Demand. Descriptions of these services may be found on the GÉANT website at [www.geant.net](http://www.geant.net).

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 can 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.

Moreover a number of user access and application services for the research community are provided by GÉANT, including eduROAM (facilitating access to wireless networks in campuses around the world), eduGAIN (providing interoperation between national digital identity federations), and eduCONF (to provide a means of certifying video conference facilities).

### **5.2.2 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.

#### **5.2.2.1 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 PATC (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 could 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. PATCS updates its curriculum yearly with following the demand from users and with 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 High Performance Computing (HPC), thus increasing their competitiveness.

#### **5.2.2.2 EGI services**

EGI provides four solutions to accelerate data-intensive research. These solutions are built through a service catalogue that has been evolving according to the user requirements

The **EGI Federated Cloud Solution** offers an infrastructure to deploy on-demand IT services. It can host datasets of public or commercial relevance, and can be flexibly expanded by integrating new providers. The infrastructure is a single, standards-based, open system to federate academic and commercial clouds from multiple providers. 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** provides a global high-throughput data analysis infrastructure, linking hundreds of independent research institutes, universities and organisations delivering computing resources and high scalability. This solution is aimed to help individual researchers, and research communities that have large scale distributed data management and computational capacity requirements. With this solution users gain access to 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 complementary 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 three access models:

- Free at point of use with grant-based allocations or MoUs: this model applies to data and resource centres which can offer long-term data and compute services for free to the RIs that are part of a national roadmap;
- Pay per use: the RI can freely purchase data and compute services from a market of providers following traditional best value for money procurement strategies;
- Annual membership fees.

The first two access models are applicable to high throughput computing and cloud: in these cases access policies depend on the service providers of choice and can vary nationally and regionally.

The annual membership fee is the current access model for the federated operations solution.

#### **5.2.2.3 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 through the innovative open source broker technology. Federation of public sector data centre resources from within the European Grid Infrastructure (EGI) with HNX will enable EGI's research communities to use HNX services within a hybrid cloud model. 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 via HNX.

The Cloud services are offered to the global research community, for both publicly-funded and commercial organizations of diverse sectors.

#### **5.2.2.4 GÉANT cloud services**

GÉANT is actively helping NRENs (National Research and Education Networks) 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.

### **5.2.3 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.

A data management policy is vital for each ESFRI project proposal and to the extent the ESRPRI project itself is generating/collecting data, it should also have a data management plan (DMP). For ESFRI projects (typically instruments / laboratories) where the users (research projects) generate the data it are the latter that need to have a DMP. The data management policy can address access rights and restrictions, long-term usability, data formats, meta-data structures, availability within, or interoperability, with existing data sharing frameworks. The DMP outlines how data will be handled during the research and after completion of the research project. A DMP 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 of the research projects 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 successful data management plans are available and mentioned in the References section.

#### ***Data infrastructures and related services***

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 at the Horizon 2020. 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 are:

- **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 centers 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** 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).

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 ERCs. 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.

### *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**”<sup>5</sup>. 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**”<sup>6</sup> 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**<sup>7</sup> and **White Paper 2013**<sup>8</sup>.

## **6 Best practices and recommendations**

The above chapters show that a number of best practices and recommendations can be defined when analysing the e-infrastructure needs of Research Infrastructures.

First, the e-Infrastructure needs and data aspects have to be fully taken into account from the beginning of the RI 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 RIs, sustainable e-Infrastructure services for enabling access to, storing, preserving and curating large amounts of data need to be in place. Also, to ensure that data will be available across borders and disciplinary domains, RIs 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).

---

<sup>5</sup> <http://zenodo.org/record/7636>

<sup>6</sup> <https://europe.rd-alliance.org/documents/publications-reports/data-harvest-how-sharing-research-data-can-lead-to-knowledge-jobs-and-growth>

<sup>7</sup> [http://e-irg.eu/documents/10920/238805/e-irg-blue\\_paper\\_on\\_data\\_management\\_v\\_final.pdf/d2de68e7-2360-42c8-8a08-e42ad92af06e](http://e-irg.eu/documents/10920/238805/e-irg-blue_paper_on_data_management_v_final.pdf/d2de68e7-2360-42c8-8a08-e42ad92af06e)

<sup>8</sup> <http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf/ce8a2253-aebd-4cbe-9a93-4709a1166214>

- 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 RI 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, see the EC document Charter for access to Research Infrastructures.
- Pay attention to the sustainability of your data, also after the end of the project.

## 7 Use Cases

Organisations and projects are successfully using e-Infrastructures. 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 RIs is also possible. Below we point to success stories in the use of e-Infrastructures by research projects.

### **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](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**

**HELIX NEBULA use**

## **8 References**

### **Information about national e-Infrastructures**

<http://knowledgebase.e-irg.eu/countries>

### **Information GEANT Association**

<http://www.geant.org/Pages/Home.aspx>

### **TERENA NREN Compendium**

<http://www.terena.org/activities/compendium/index.php?showyear=2013>

### **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://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.pdf](http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf) DPHEP, <http://www.dphep.org/>

See also the 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](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

## 9 Glossary

Cloud	distributed computing and storage infrastructure able to provide services on different levels of abstractions (IaaS, SaaS, PaaS)
EGI	European Grid Initiative
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
NGI	National Grid Initiative
NREN	National Research and Education Network
OpenAIRE	Open Access Infrastructure for Research in Europe
PRACE	Partnership for Advanced Computing in Europe
RI	Research Infrastructure
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