

CONTENTS

118	NETWORKING INFRASTRUCTURES
118	CURRENT STATUS
120	COMPUTING, DATA AND CLOUD INFRASTRUCTURES
120	CURRENT STATUS
122	GAPS, CHALLENGES AND FUTURE NEEDS

PART2

LANDSCAPE ANALYSIS - SECTION1

DATA, COMPUTING AND DIGITAL RESEARCH INFRASTRUCTURES

DATA, COMPUTING AND DIGITAL RESEARCH INFRASTRUCTURES

In research, as in all fields of society, Information and Communications Technology (ICT) has become a key enabling factor for progress. ICT is also changing the *modus operandi* of research by providing new possibilities for geographically distributed collaboration and sharing. Data-driven science, as well as more and more open access to data and scientific results, is transforming not only how research is conducted, but its overall reach.

Today, all large-scale Research Infrastructures are already dependent on ICT resources. This dependence has increased the need to find synergies and to develop ways to tackle the ICT challenges at a generic level, providing effective and cost efficient services that can be of wide and general use. The pan-European e-Infrastructures for networking, high-performance computing (supercomputing) and high-throughput computing (clusters built from more commodity-type hardware) are already well-established and provide production services used by international research and Research Infrastructures projects. Also, data and cloud infrastructures are developing fast and consolidation and integration of such initiatives is taking place, partly inspired by the description of the European e-Infrastructure *Commons* as proposed by the e-Infrastructure Reflection Group (e-IRG). The European Commission enforces the collaboration between researchers, Research Infrastructures and e-Infrastructures by the European Open Science Cloud (EOSC) declaration and the

Electronic Data Infrastructure (EDI) definition as a combination of very fast networking and exascale computing.

The e-Infrastructure landscape described below paves the way for common solutions for shared needs and requirements. Having a general, common layer of supporting e-Infrastructures – *horizontal* e-Infrastructures – also allows for a refocus on science for the disciplinary Research Infrastructures. *Horizontal* e-Infrastructures shared by thematic infrastructures are an important facilitator of cross-disciplinary work, thereby enabling the study of fundamentally new research questions. Here, it should be noted that the development of ICT is very fast and further innovation and development of e-Infrastructure services of all types is essential to make sure that the needs of the European research and Research Infrastructures communities can also be tackled in the future.

Beyond e-Infrastructures and digital devices for network, computation and data man-

agement, it is also of a major importance to develop specific Research Infrastructures in the domain of computer sciences, supporting the experimentation of disruptive systems including e-Infrastructures, hard-middle-sofwares, protocols, computing and cybersecurity issues. This is best done as a co-design effort between all stakeholders where new needs of the researchers lead the way to innovation efforts.

The European e-Infrastructure landscape includes a networking infrastructure, computing facilities and data infrastructures. The European e-Infrastructure ecosystem fully includes national, regional and institutional e-Infrastructures. It should be noted that the pan-European e-Infrastructure services are often being provided by national e-Infrastructures in a collaborative setting, and the European initiatives are dependent on the existence of strong, persistent and synergic national e-Infrastructure nodes.

In the subfields a brief introduction of the major pan-European *horizontal* e-Infrastructure initiatives is given and some examples of services provided are listed. The EOSC and EDI initiatives are highlighted in the corresponding sections. A more complete account of available services can be found in the e-IRG Guidelines Document 2017¹.

The subfields in the e-Infrastructures domain are **NETWORKING INFRASTRUCTURES** and **COMPUTING, DATA AND CLOUD INFRASTRUCTURES**.

1. Guide to e-Infrastructure requirements for European Research Infrastructures, e-IRG support document, 2017 <http://e-irg.eu/documents/10920/363494/2017-Supportdocument.pdf>

NETWORKING INFRASTRUCTURES

II CURRENT STATUS

GÉANT² manages the pan-European research and education network which links together and offers transnational access to RIs and research resources by providing interconnectivity between National Research and Education Networks (NRENs) across 43 European countries. The NRENs connect universities, research institutes, and sometimes other public institutions in each country. The access to NREN resources is managed nationally and the policy differs slightly from country to country. In addition to pan-European connectivity, the GÉANT network has international connections to a large set of partner networks worldwide, enabling international collaboration on research and education. Most large-scale Research Infrastructures can connect to the local NREN and thus access GÉANT enabling worldwide communications. Projects can also work with their NREN and GÉANT for international point to point links to connect parts of the Research Infrastructures that are distributed over Europe or beyond. If the project or infrastructure is distributed across national boundaries, GÉANT can help coordinate with the relevant local NRENs and advise on appropriate technical solutions. GÉANT also provides important services for researchers, such as innovation test beds.

GÉANT delivers a range of networking services at the international level. Most of these services match those offered at national level by the NRENs. The GÉANT NREN Compendium³ gives an overview of all NRENs, their services and the contact information. The connectivity delivered by GÉANT is supported by a comprehensive range of network monitoring and management services for e.g. optimizing network performance and continuous monitoring. Users can also benefit from the range of GÉANT network monitoring, security and support services employed by NRENs to assure optimum performance for projects and institutions.

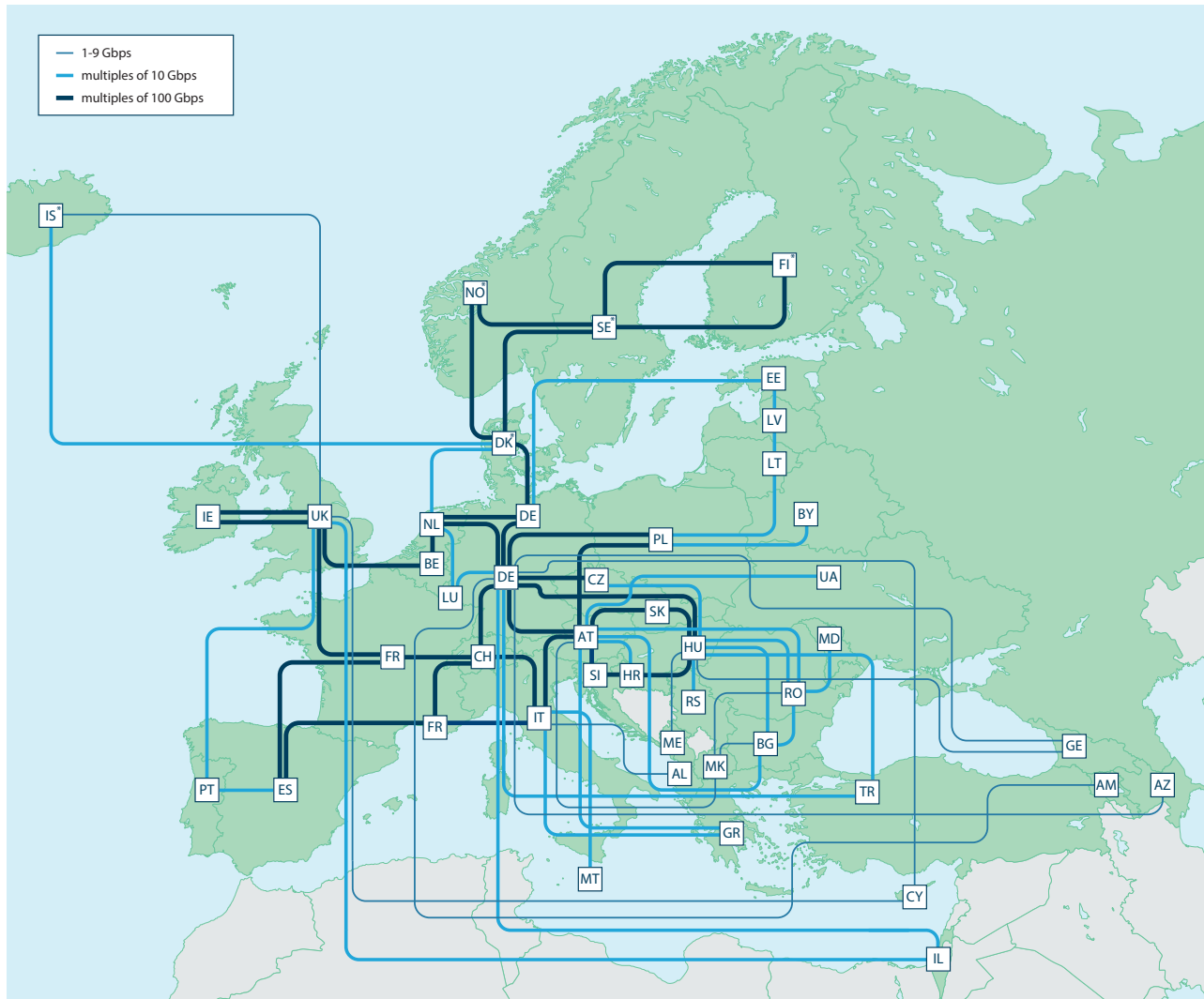
Leading research is today often conducted in a highly distributed and mobile environment where researchers freely collaborate across boundaries. The research communities need to manage access to their services from participants in many organisations and individual researchers need to easily and securely access multiple tools, services and datasets. Trust and identity therefore take up a pivotal position in the e-Infrastructure ecosystem. Here, federated authorization and authentication services simplify access to inter-organisational resources, allowing controlled and secure access. By forming a layer connecting the power of the network with computing, data and cloud infrastructures, such services enable safe and secure research throughout Europe and beyond. In this context, a number of services for the research community are provided to the research community by GÉANT. These services include eduroam that facilitates access to wireless networks in campuses around the world and eduGAIN that provides a framework for interoperation between digital identity federations.

2. GÉANT
<https://www.geant.org/>

3. GÉANT NREN Compendium
<https://compendium.geant.org/#/>



GÉANT's pan-European **research and education network** interconnects Europe's National Research and Education Networks (NRENs). Together we connect over **50 million users** at 10,000 institutions across Europe.



GÉANT's pan-European network is funded by the GÉANT Project (GN4-2), which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 731122. Connectivity to the Eastern Partnership countries (AM, AZ, BY, GE, MD, UA) is provided through the EaPConnect project, with 95% funding from the European Commission DG NEAR under grant agreement 2015-356353. The map shows topology as at January 2018. The GN4-2 and EaPConnect partners are listed below.

RASH AL Albania	Belnet BE Belgium	Cyprus CY Cyprus	EENet EE Estonia	RENA GE Georgia	HEAnet IE Ireland	LITHNET LT Lithuania	OMREN ME Montenegro	NORDUnet NO Norway*	AMRES RS Serbia	TURKNET TR Turkey
Asnet AM Armenia	BULNET BG Bulgaria	cesnet CZ Czech Republic	IRIS ES Spain	grnet GR Greece	ISRI IL Israel	LUXNET LU Luxembourg	MARnet MK F.Y.R. Macedonia	PSNC PL Poland	NORDUnet SE Sweden*	Jisc UK United Kingdom
oconet AT Austria	BASNET BY Belarus	DFN DE Germany	NORDUnet FI Finland*	CARNET HR Croatia	NORDUnet IS Iceland*	LITHNET LV Latvia	Maltanet MT Malta	FCT PT Portugal	arXiv SI Slovenia	UkrNet UA Ukraine
AZERBAIJAN AZ Azerbaijan	SWITCH CH Switzerland	NORDUnet DK Denmark*	RENATER FR France	GARR HU Hungary	GARR IT Italy	MOLNET MD Moldova	EURONET NL Netherlands	ROUNET RO Romania	sanet SK Slovakia	

*Connections between these countries are part of NORDUnet (the Nordic regional network)

FIGURE 1.

This map has been downloaded from the GÉANT website. Last update: January 2018
https://www.geant.org/Resources/Documents/G%C3%89ANT_Topology_Map_January_2018.pdf

COMPUTING, DATA AND CLOUD INFRASTRUCTURES

II CURRENT STATUS

The most well-established pan-European computing infrastructures are EGI⁴ in the area of high-throughput computing and cloud infrastructures, and the **ESFRI Landmark PRACE** (Partnership for Advanced Computing in Europe) in the area of High-Performance Computing (HPC) infrastructure. EUDAT⁵ and OpenAIRE are the initiatives that focus on data Infrastructures. Helix Nebula, GÉANT and EGI offer cloud services.

These e-Infrastructures represent the core of the European e-Infrastructures however this landscape may change in the coming years under the influence of the EOSC and EDI initiatives of the EC. The EOSC Declaration⁶ states that the EOSC infrastructure will be developed as a data infrastructure *Commons* serving the needs of scientists. It should provide both common functions and localised services delegated to community level. Indeed, the EOSC will federate existing resources across national data centres, European e-Infrastructures and Research Infrastructures; service provision will be based on local- to-central subsidiarity – e.g. national and disciplinary nodes connected to nodes of pan-European level; it will top-up mature capacity through the acquisition of resources at pan-European level by EOSC operators, to serve a wider number of researchers in Europe. Users should contribute to define the main common functionalities needed by their own community. A continuous dialogue to build trust and agreements among funders, users and service providers is necessary for sustainability. The EC has included a number of calls for proposals to build the EOSC in its Research Infrastructures and e-Infrastructures Work Programmes 2016-2017 and 2018-2020. It is anticipated that in the framework of these calls current e-Infrastructures, as mentioned here, will cooperate ever more closely, such as EGI and EUDAT in the EOSCHub project.

4. EGI
<https://www.egi.eu/>

5. EUROPEAN DATA INITIATIVE
<https://www.eudat.eu/european-data-initiative>

6. EOSC Declaration
<https://ec.europa.eu/research/openscience/pdf/eosc-declaration.pdf>

The supercomputing landscape will be enhanced in the coming years through the EuroHPC initiative. EuroHPC is part of EDI and aims at providing exascale computing in Europe. On 23rd March 2017, seven European countries signed an agreement to start a European HPC programme that will eventually lead to European exascale supercomputers called EuroHPC. Meanwhile, eight more MS have joined the declaration. Those Member States agreed to work together and with the European Commission in the context of a multi-government agreement called EuroHPC for acquiring and deploying by 2022/2023 a pan-European integrated pre-exascale supercomputing and a data infrastructure that will support data-intensive advanced applications and services. It is a response to the surging demand from scientists, industry and the public sector for access to leading-edge computing capacity to cope with vast amounts of data produced in almost all scientific and engineering domains. The EuroHPC joint undertaking will provide EU level coordination and adequate financial resources to support the development and procurement of such infrastructure. This infrastructure will be accessible to public and private users for research purposes; paying services to industry may also be provided (under conditions TBD). The EuroHPC has proposed, by the end of 2017, a legal instrument that provides a procurement framework for the exascale supercomputing and data infrastructure.

EGI is an international collaboration that federates the digital capabilities, resources and expertise of national and international research communities in Europe and worldwide. EGI's main goal is to empower researchers from all disciplines to collaborate and to carry out data- and compute-intensive science and innovation. EGI is coordinated by the EGI Foundation and has participants from national representatives (NGIs), EIROforums and ERICs. EGI provides open solutions offered through a service catalogue that has been evolving for many years. The EGI Federated Cloud Solution offers a standards-based and open infrastructure to deploy on-demand IT services that can manage and process datasets of public or commercial relevance, and can be flexibly expanded by integrating new providers. This is complemented by the EGI High Throughput Computing Solution which provides a global high-throughput data analysis infrastructure,

linking a large number of independent organisations and delivering computing resources and high scalability. The EGI Federated Operations Solution provides processes and tools to federate and manage distributed ICT capabilities. The EGI Community-Driven Innovation & Support Solution provides the processes, framework and experts so that research communities can co-create the new capabilities or adapt their existing applications or platforms for compute- or data- intensive science on EGI. Access to EGIs externally provided resources is provided through three different access modes: using free grant-based allocations, pay per use, and annual membership fees. The first two modes are applicable to the high throughput computing and cloud solutions and the policies depend on the service providers of choice and can vary nationally and regionally.

The **ESFRI Landmark PRACE** offers access to world-class high-performance capability computing facilities and services. PRACE is managed by the PRACE AISBL and is governed by governmental representative organisations. 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 scientific peer-review and open R&D. The PRACE 2 epoch that has been launched in the beginning of 2017 welcomes a new host hence making more computing cycles available to the research community and guarantees PRACE sustainability until 2020. PRACE is only briefly presented here, further details can be found in the dedicated card in the ESFRI Landmarks section of **Part3**.

In some countries, the national representatives are the same for EGI and PRACE. Both EGI and PRACE have already established contacts with consortia that operate or prepare European large-scale Research Infrastructures to understand needs and find out how these matches with available resources and existing policies.

The amount of digital information is growing rapidly. Large-scale Research Infrastructures, such as the initiatives on the ESFRI roadmap, produce and are dependent on a rapidly increasing amount of data. The importance of data management has emerged as a key element in many large-scale Research Infrastructures.

tures projects. It is recognised that specific efforts are needed for making data discoverable and reusable, but data sharing preparedness even within disciplines still differs a lot. The data infrastructures developed by disciplinary Research Infrastructures are often, for natural reasons, customised for the concerned project or research discipline domain and not primarily aimed at use beyond the project or discipline borders. In fact, several of the existing European large-scale Research Infrastructures could be classified as disciplinary e-Infrastructures focussing on disciplinary interoperability and access to data. Several ESFRI cluster projects have been studying similarities between the data needs of sets of ESFRI Research Infrastructures, considering common data standards and formats, data storage facilities, integrated access and discovery, data curation, privacy and security, service discovery and service market places.

For research and society to take full benefit of the major investments in Research Infrastructures and research, the data needs to be made openly and easily available for researchers, over wide spans of fields, in sustainable settings. Also, the data needs to be managed, stored and preserved in a cost-efficient way and the access to the data across borders and domain boundaries must be secured. To fully exploit the underlying potential value in the rapidly increasing amount of research data, interoperability between data infrastructures at all levels is becoming crucial. Efforts have been made to attain a common understanding on the realisation of an ecosystem of data infrastructures and related services, including producing a set of joint recommendations by ESFRI and e-IRG⁷. Many disciplines work at the European and international level to define the discipline-specific aspects of their data infrastructure, which then should be interfaced with the more generic data infrastructure components to provide cross-field interoperability.

Much effort is today going into the definition and development of common or interoperable data formats and metadata, which is necessary to fulfil the general requirement to provide data following the FAIR – Findable, Accessible, Interoperable and Reusable – principles. This requires significant engagement and work

from scientific communities at disciplinary level as a starting point to define standards and provide reusable data, as well as data management services to enable data interoperability and sharing, aiming at the realisation of an ecosystem with the appropriate technical and social channels for openly sharing of data at a multidisciplinary and global level. Here, an active role is played by the Research Data Alliance (RDA)⁸ initiative, a bottom-up organisation with constituents in different regions – such as RDA Europe – and countries. The goal of RDA is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, and the work is performed in Working and Interest Groups which tackle diverse sociological and technological aspects of research data sharing. At the European level, data infrastructures are not yet as well-established as the basic networking and computing infrastructures. However, significant steps have been made in the areas of basic data services (such as storage and replication) through the EUDAT projects and access to publications and other research results through the OpenAIRE projects.

EUDAT is the largest pan-European data infrastructure initiative and has now taken the necessary steps to move towards a sustainable 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 services include a secure and trusted data exchange service, a data management and replication service, a service to ship large amounts of research data between EUDAT data nodes and workspace areas of high-performance computing systems and 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.

OpenAIRE enables researchers to deposit research publications and data into Open Access repositories and provides support to researchers at the national, institutional and local level to guide them on how to publish in Open Access (OA) and how to manage the long tail of science data within the institution environment. If researchers have no access to an in-

stitutional or a subject repository, Zenodo⁹, hosted by CERN, enables them to deposit their articles, research data and software. Zenodo 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 ERC. Zenodo has also been extended with important features that improve data sharing, such as the creation of persistent identifiers for articles, research data and software.

The Helix Nebula initiative is providing a public-private partnership 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. A series of cloud service procurement actions, including joint pre-commercial procurement co-funded by the EC, are using the hybrid public-private cloud model to federate e-Infrastructures with commercial cloud services into a common platform delivering services on a pay per use basis.

7. Summary of Policy Recommendations Drawn from the e-IRG Blue Paper on Data, e-IRG Blue Paper, 2013 <http://e-irg.eu/documents/10920/238805/BP-summary-policy-130227.pdf>

8. Research Data Alliance <https://www.rd-alliance.org/>

9. Zenodo <https://zenodo.org/record/7636#.W0yQfgUzaUk>

▶ GAPS, CHALLENGES AND FUTURE NEEDS

The e-IRG has identified the need for a more coherent e-Infrastructure landscape in Europe, in particular in its 2013 White Paper¹⁰. By now, this notion of a European e-Infrastructure *Commons* framework has been widely accepted and several steps have been taken towards its implementation, including the realisation of EOSC. The e-Infrastructure *Commons* framework has acted as the solid basis for designing the EOSC and its implementation program, already containing most of the ingredients needed for an integrated European platform for Open Science.

The e-Infrastructure *Commons* is the framework for an easy and cost-effective shared use of distributed electronic resources for research and innovation across Europe and beyond. An essential feature of the *Commons* is the provisioning of a clearly defined, comprehensive, interoperable and sustained set of services, provisioned by several e-Infrastructure providers, both public and commercial, to fulfil specific needs of the users. This set should be constantly evolving to adapt to changing user needs, complete in the sense that the needs of all relevant user communities are served and minimal in the sense that all services are explicitly motivated by user needs and that any overlap of services are thoroughly motivated. The *Commons* has three distinct elements:

- a platform for coordination of the services building the *Commons*, with a central role for European research, innovation and Research Infrastructures communities;
- provisioning of sustainable and interoperable e-Infrastructure services within the *Commons*, promoting a flexible and open approach where user communities are empowered to select the services that fulfil their requirements;
- implementation of innovation projects providing the constant evolution of e-Infrastructures needed to meet the rapidly evolving needs of user communities.

In summary, the ultimate vision of the *Commons* is to reach integration and interoperability in the area of e-Infrastructure services, within and between Member States, and at the European level and globally. It is the mission of e-IRG to support this vision through supporting a coherent, innovative and strategic European e-Infrastructure policy making and the development of convergent and sustainable e-Infrastructure services. This e-Infrastructure *Commons* is also a solid basis for building the EOSC already containing most of the ingredients needed for an integrated European platform for Open Science.

In its Roadmap 2016 document¹¹, the e-IRG provides the following key recommendations to attain this *Commons* or EOSC:

- 1) Research infrastructures and research communities should reinforce their efforts to:
 - elaborate on and drive their e-Infrastructure needs;
 - participate in the innovation of e-Infrastructure services;
 - contribute to standards and take care of their data.
- 2) e-Infrastructure providers should further increase their efforts to work closely together to fulfil the often complex user needs in a seamless way.
- 3) National governments and funding agencies should reinforce their efforts to:
 - embrace e-Infrastructure coordination at the national level and build strong national e-Infrastructure building blocks, enabling coherent and efficient participation in European efforts;
 - together analyse and evaluate their national e-Infrastructure funding and governance mechanisms, identify best practices, and provide input to the development of the European e-Infrastructure landscape.

4) The European Commission should provide strong incentives for cross-platform innovations and further support the coordination and consolidation of e-Infrastructure service development and provisioning at the national and the European level.

10. e-IRG White Paper
<http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf>

11. e-IRG Roadmap 2016 document
http://e-irg.eu/documents/10920/12353/Roadmap*2016.pdf

