

# CONTENTS

<b>42</b>	<b>CURRENT STATUS</b>
43	NETWORKING AND OTHER SERVICES
44	DATA INFRASTRUCTURES
45	COMPUTING INFRASTRUCTURES
46	THEMATIC e-INFRASTRUCTURES
<b>47</b>	<b>GAPS, CHALLENGES AND FUTURE NEEDS</b>

PART2

LANDSCAPE ANALYSIS - SECTION1

# DATA, COMPUTING & DIGITAL RESEARCH INFRASTRUCTURES

# DATA, COMPUTING & DIGITAL RESEARCH INFRASTRUCTURES

A Digital Infrastructure (DI) is broadly defined as a set of information and communication technology components that are the foundation of ICT-services. These include typically physical components - computer and networking hardware and facilities - but also various software and network components.

The term e-Infrastructure is commonly used to indicate advanced fully integrated communication and information processing services, which provide transparent, easy, cheap and secure access to all types of distributed resources - computers, databases, heavy research instruments.

Digital Infrastructures are expected to boost research, growth, innovation and job creation, and it is clear that education of digital scientists and practitioners are a priority for Europe as this can effectively give people the knowledge, skills and competences to use and benefit from scientific data. Recent discussions around 5G and COVID-19 contact tracing amplify the need for European sovereignty when it comes to Digital Infrastructures and the handling of data. It is important to recognise that Digital Infrastructure control and data regulation are complementary and can be combined in various ways<sup>1</sup>.

The European e-Infrastructure landscape includes Networking, Computing, and Data Infrastructures, on the national, regional and institutional level. The e-Infrastructure services at the European level are often being provided by federating national e-Infrastructures in a collaborative setting, and the European initiatives are therefore dependent on the existence of strong and

coherent national e-Infrastructure nodes<sup>2</sup>. To describe this ecosystem, e-IRG has introduced the concept of the e-Infrastructure Commons in its White Paper 2013<sup>3</sup>. Building on this and other notions the concept of the European Open Science Cloud (EOSC) has emerged.

The **EUROPEAN OPEN SCIENCE CLOUD** is an environment for hosting and processing research data to support European research. The EOSC gained rapidly impetus and attention with the definition of an implementation roadmap in March 2018<sup>4</sup>.

The European Open Science Cloud intends to offer to the researchers a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing Data Infrastructures.

EOSC is being co-created in a series of projects funded by the European Commission and initiatives from Member States (MS) and Associated Countries (AC). The contribution of MS and AC will constitute the bulk of the resources that will be made available by the contributing organisations and thus their role is of fundamental importance also in the governance of the EOSC. EU coun-

tries and countries associated with Horizon 2020, represented in the EOSC Governance Board, agreed unanimously to run the EOSC as a co-programmed European Partnership under Horizon Europe from 2021.

The EOSC Association<sup>5</sup> was established as International non-profit Association under Belgian Law (AISBL) law on 29<sup>th</sup> July 2020 with four founding members: GÉANT, CE-SAER, CSIC and GARR. Members and Observers of the EOSC Association include research funders, research performing organisations, Research Infrastructures, data service providers and others. As of today the activities of the Working Groups of the EOSC Executive Board have produced several important results<sup>6</sup> one of which is the document *Solutions for a Sustainable EOSC - A FAIR Lady report* from the EOSC Sustainability Working Group, which explored possible means for sustaining the EOSC beyond its initial phase. The launch of the European Partnership for the EOSC, in close cooperation with the Member States and the respective research communities was on the 23<sup>rd</sup> of June 2021. The new EOSC European Partnership will ensure, until at least the end of 2030, a coordinated approach from the European Commission and the stakeholders in investments and initiatives in the EOSC ecosystem with the help of the Member States and Associated Countries. An EU investment of almost € 500 million and an in-kind contribution of the partners of also € 500 million are foreseen in the period 2021-2027. The aim is to improve the storing, sharing and reusing of research data across borders and scientific disciplines<sup>7</sup>.

1. [New report on European Digital Infrastructure and Data Sovereignty. A policy perspective \(2020\)](https://www.eitdigital.eu/newsroom/news/archive/article/new-report-on-european-digital-infrastructure-and-data-sovereignty/)  
<https://www.eitdigital.eu/newsroom/news/archive/article/new-report-on-european-digital-infrastructure-and-data-sovereignty/>

2. [National Nodes - Getting organised: how far are we?](http://e-irg.eu/catalogue/eirg-1006)  
e-Infrastructure Reflection Group (2020)  
<http://e-irg.eu/catalogue/eirg-1006>

3. [e-IRG White Paper 2013, e-Infrastructure Reflection Group \(2013\)](http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf)  
<http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf>

4. [Implementation Roadmap for the European Open Science Cloud. Commission Staff Working Document \(2018\)](https://ec.europa.eu/transparency/regdoc/rep/10102/2018/EN/SWD-2018-83-FI-EN-MAIN-PART-1/PDF)  
<https://ec.europa.eu/transparency/regdoc/rep/10102/2018/EN/SWD-2018-83-FI-EN-MAIN-PART-1/PDF>

5. [EOSC Association](https://www.eosc.eu/)  
<https://www.eosc.eu/>

6. [Working Groups of the EOSC Executive Board](https://www.eoscsecretariat.eu/eosc-working-groups)  
<https://www.eoscsecretariat.eu/eosc-working-groups>

7. [Launch ceremony of the EOSC European Partnership](https://ec.europa.eu/info/news/launch-ceremony-eosc-european-partnership-during-ri-days-2021-2021-jun-23_en)  
[https://ec.europa.eu/info/news/launch-ceremony-eosc-european-partnership-during-ri-days-2021-2021-jun-23\\_en](https://ec.europa.eu/info/news/launch-ceremony-eosc-european-partnership-during-ri-days-2021-2021-jun-23_en)

Since the beginning of the year 2019, five ESFRI Cluster projects – PaNOSC, EN-VRI-FAIR, EOSC-Life, ESCAPE, SSHOC – have been launched to link to the European Open Science Cloud. The five ESFRI Cluster projects aim together to implement interfaces, to integrate computer and data management solutions, to create cross-border and open cooperation spaces and to promote clouds via the EOSC portal for a larger user community. As stakeholders of EOSC these projects were invited to contribute to the development and implementation process. The overall expectation of these projects is that EOSC will enable the accessibility and re-use of research data, increase scientific value of research data, and deliver an interoperable environment of Data Infrastructures. The projects expect EOSC will bring the added values of the infrastructure for sustainable use of research data and a virtual research environment enabling real-time collaboration between researchers using FAIR data<sup>8</sup>.



## CURRENT STATUS

e-Infrastructures<sup>9</sup> address the needs of European researchers for digital services in terms of networking, computing and data management, and foster the emergence of open science as an essential block of the European Research Area (ERA). Federated, national Infrastructures and European initiatives will benefit scientific communities by providing trusted and open environments to store, share and re-use scientific data and results, as well as benefit from fast connectivity, high-capacity cloud solutions, and supercomputing capability system.

Throughout the MS and AC, a large variety of data processing services are available: from local, regional, and national services to international services. Large, international scientific collaborations have often created their own e-Infrastructure ecosystem, for example, the Worldwide LHC Computing Grid (WLCG), the highly distributed data processing approach for CERN's LHC experiments. Historically, two general classes of computation have provided data processing: High-Throughput Computing (HTC) and High-Performance Computing (HPC). HTC systems involve running many independent tasks that require a large amount of computing power and are optimised for large data processing tasks. HPC is commonly used to describe super-computing facilities, which process data in parallel and are optimised for a maximum number of computing operations per second. Although communities and their use cases could be generally assigned to one of the two computing models until around 2010, in recent years, more and more cases of heterogeneous use cases have emerged, which require a mix of both, high-data throughput and large number of computing operations per second, and thus, demand heterogeneous systems. The separation of the different types of computing e-Infrastructures is at least in part due to European funding strategies and organisations.

Based on a partnership model, EGI and EUDAT co-ordinate at an international level significant HTC and data services, whereas HPC centres join the **ESFRI Landmark PRACE** partnership initiative and participate in EuroHPC. The mission of PRACE is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE seeks to realise this mission by offering world-class computing and data management resources and services through a peer-review process.

The **ESFRI Project SLICES** is a distributed e-Infrastructure that focuses mainly on cloud and edge computing, Internet of Things (IoT) and networking/future internet. Traditionally, the e-Infrastructures focussed on centralised High-Performance Computing, distributed High-Throughput Computing, storage or network but it is true that dedicated e-Infrastructure on cloud and edge computing, IoT and networking has been missing in the European Roadmap of Research Infrastructures. As the clear difference between computing, network and storage is vanishing, the infrastructure SLICES in the Roadmap 2021 covers the gap. Providing the research and engineering community with a fully controllable, programmable virtualized Digital Infrastructure test platform distinguishes this infrastructure from more traditional/operational infrastructures as well. SLICES will allow academics and industry to experiment and test future, possibly long-term and disruptive DIs, essential for European research: a holistic and comprehensive approach whereby all computing, networking, storage, and IoT resources can be combined to continuously design, experiment, operate, and automate DIs full life cycle management, providing a playground for research on Future Internet and distributed systems.

The **ESFRI Project SoBigData++** (henceforth SBD++) aims to establish a European infrastructure of big data and social data mining, using new methods and implementing it in differ-

8. ESFRI cluster projects position papers (January 2020)  
[https://www.fairsfair.eu/sites/default/files/ESFRI\\_clusters\\_position\\_on\\_EOSC\\_jan\\_2020\\_v1.pdf](https://www.fairsfair.eu/sites/default/files/ESFRI_clusters_position_on_EOSC_jan_2020_v1.pdf)

9. Landscape of EOSC-related infrastructures and initiatives. Report from the EOSC Executive Board Working Group (WG) landscape (2020)  
<https://op.europa.eu/en/publication-detail/-/publication/cbb40bf3-f6fb-11ea-991b-01aa75ed71a1>

ent fields of data analysis. This is in line with current scientific trends in machine learning and data science to promote ethically sound and open research in large datasets that democratizes the value of data science. SBD++ is expecting to become the leading RI for realizing large-scale social mining experiments.

The **ESFRI Project EBRAINS** is defined as the *one-stop-shop* that is offering scientists and developers the most advanced tools and services for brain research. Human Brain Project (HBP), which is one of the FET flagship projects, is the developer and the provider of EBRAINS. As HBP has the internal means to create a self-contained structure, it is a challenge to create an outward looking environment and the questions that are raised show much this is already the case and how much effort would it need to serve groups that are not part of the original HBP project. The overall Digit Landscape is represented in **Figure 1**.

## NETWORKING AND OTHER SERVICES

Today, each European country has a National Research and Education Network (NREN), connecting research and higher education institutions with high-performance networks, and offering a range of related services.

In terms of organisation and funding, the European NRENs are diverse. Some are funded directly from government budget; others are funded by their connected institutions. Some are part of large organisations managing a variety of national e-Infrastructures, while others are smaller organisations focussing on just the network. Nevertheless, they have important similarities. All NRENs offer high-performance networks suited to the special needs of research and education, with the headroom required for the bursts in traffic which are unique to research and large instruments, and the capability to serve research collaborations like ESFRI's with specialised network support.

Additionally, all European NRENs offer critical access and identity services such as eduroam and eduGAIN. These trust and identity services make up the foundation of services that allow secure access to research data, authentication to shared resources, and support for mobility and collaboration. Many NRENs also offer storage services, computing services, and a range of security services.

Together, the NRENs have formed the **GÉANT** Association, an organisation for European collaboration in research networks and the operator of the pan-European GÉANT network, with connectivity to other world regions. With support from the EC during decades of Framework Partnerships, the GÉANT network has been developed into a world-leading network, ensuring world-class connectivity to all European countries and making Europe a leading actor in global research networking and e-Infrastructures.

Through its integrated catalogue of connectivity, collaboration and identity services, GÉANT provide users with highly reliable, unconstrained access to computing, analysis, storage, applications and other resources, to ensure that Europe remains at the forefront of research. GÉANT interconnects 38 NREN partners, and it is the largest and most advanced Research & Education (R&E) network in the world. GÉANT connects over 50 million users at more than 10,000 institutions across Europe and supports all scientific disciplines. The backbone network operates at speeds of up to 500 Gbps and reaches over 100 national networks worldwide. Since its establishment over 20 years ago, the GÉANT network has progressively developed to ensure that European researchers lead international and global collaboration. Over 1,000 terabytes of data are transferred via the GÉANT IP backbone every day.

More than just an Infrastructure for e-science, it stands as a positive example of European integration and collaboration. GÉANT develops and delivers advanced networks and associated e-Infrastructure services. It supports open innovation, collaboration and knowledge sharing amongst its members, partners and the wider research and education networking community. With more than 40 partners and associates across Europe and a multi-million euro budget, GÉANT has met the challenge of complex international project management. GÉANT also provides consultancy on network-related projects. GÉANT has national members (one per state) and representative members (represent at least two legal entities of different countries), associate (no voting rights) and the possibility of establishing working committees. According to the most recent report: more than 80% of the universities are connected to GÉANT, with 86% of all university-level students serviced in those 40 countries; that is, a total of 25 million university students. The GÉANT network reaches in excess of 50 million users involved in Research & Education in the region. GÉANT network also offers connectivity to other world regions (AfricaConnect2, CAREN, EUMEDConnect3, EaP-Connect, TANDEM and others).

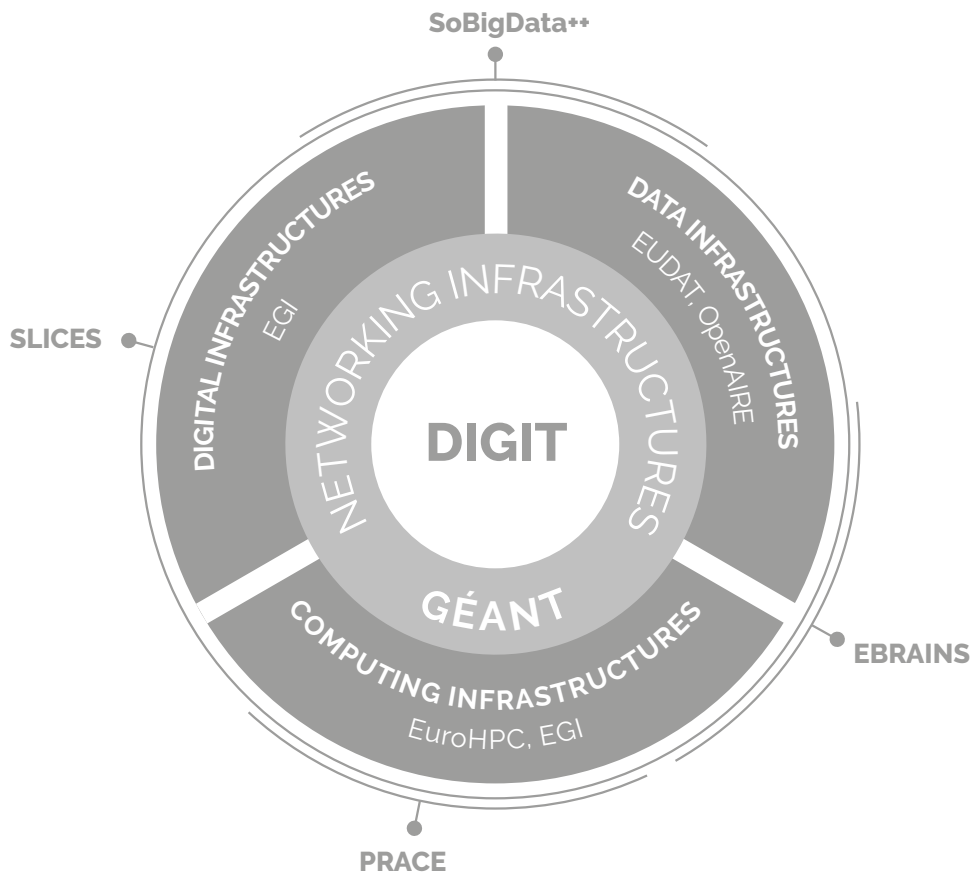


FIGURE 1.

The Landscape of the Data, Computing & Digital Research Infrastructures domain

## DATA INFRASTRUCTURES

According to the Open Data Institute definition, "Data infrastructures consist of data assets supported by people, processes and technology". In the context of this report, we consider Data Infrastructures the technical and human infrastructures, which support management and sharing of research data. The Re3Data project provides a global research data repository registry.

**EUDAT** is a Collaborative Data Infrastructure (CDI), which manages data spanning from European research data centres and community data repositories. EUDAT aims to support sharing and preserving data across borders and disciplines. European researchers and practitioners from any research discipline can preserve, find, access, and process data in a trusted environment. EUDAT offers heterogeneous research data management services and storage

resources, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network distributed across 15 European countries. Data is stored alongside some of Europe's most powerful supercomputers. One of EUDAT's main ambitions is to bridge the gap between RIs and e-Infrastructures through an active engagement strategy, using the communities that are in the consortium as EUDAT beacons and integrates others through innovative partnerships. EUDAT offers common data services, supporting multiple research communities as well as individuals, through a network of 36 European organisations. Its main services are the following: B2DROP, B2SHARE, B2SAFE, B2STAGE, B2FIND, B2HANDLE and B2ACCESS. EUDAT has a dual governance structure. As a EU-funded project, it operates through the respective bodies found in most EU projects, i.e. as defined by its Consortium Agreement. As an e-Infrastructure that provides a set of common data services, it operates on the basis of the EUDAT CDI.

Generic and thematic service providers may join the EUDAT CDI network by signing a specific collaboration agreement.

Scholarly communication initiatives and services are a relevant component of the current landscape, especially for the long tail of science. These initiatives originated from the movement to provide open access to publications, but are now applying open access principles to data (e.g. FAIR data) and other types of research products. **OpenAIRE** is a key initiative in this area having started as a supporting facility for Open Access (OA) policy of FP7 and H2020, and developing a set of mechanisms to implement and monitor open science in Europe. Services, which OpenAIRE can provide within the EOSC, are:

- a recognised national network of 35 nodes (National Open Access Desks), which are expert organisations offering local support, training and policy alignment on Open Access and Research Data Management (RDM);
- a suite of standards, the OpenAIRE Guidelines for Content providers, and services to allow content providers to make publications, data, software to share them in EOSC following open and FAIR principles (more than 1000 already registered);
- a set of services to help researchers do open science;
- Zenodo – a catch-all repository;
- Argos – an actionable DMP service linked out of the box to EU and national infrastructures;
- Amnesia – an anonymisation tool;
- the OpenAIRE Research Graph, a global contextual catalogue of research results linked together which is the basis for intelligent, AI-based discovery;
- the Open Science Observatory to monitor different aspects of open science in Europe.

## COMPUTING INFRASTRUCTURES

Computing Infrastructures typically include High-Performance Computing optimised for high memory and CPU intensive tasks and High-Throughput Computing optimised for tasks which can be divided into subtasks which distributed across multiple servers; however, Infrastructures for more specialised computing architectures also exist (e.g. GPU clusters).

At the European level, there are two significant infrastructures supporting HPC: the **EuroHPC JU** (EuroHPC Joint Undertaking, JU) and the **ESFRI Landmark PRACE**.

The EuroHPC Joint Undertaking has acquired pre-exascale and petascale supercomputers (the EuroHPC supercomputers) which will be located at and operated by supercomputing centres (Hosting Entities) in the Union. Once these systems come online, the Joint Undertaking will manage the Union's access time – from 35% up to 50% of their total capacity – of these supercomputers. From April 2021, access time will be allocated to European scientific, industrial and public sector users, matching their demanding application requirements, according to the principles stated in the EuroHPC JU Council Regulation and the JU's Access Policy.

The EuroHPC Joint Undertaking was established to enable the co-ordination of efforts and the sharing of resources at European level with the objective of deploying a world-class High-Performance Computing Infrastructure and a competitive innovation ecosystem in supercomputing technologies, applications and skills in Europe. EuroHPC JU will permit the EU and participating countries to co-ordinate their efforts and share resources with the objective of deploying in Europe a world-class supercomputing Infrastructure and a competitive innovation ecosystem in supercomputing technologies, applications and skills.

The Members of the Joint Undertaking are the following:

- the European Union, represented by the Commission;
- Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Montenegro, the Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and Turkey;
- the European Technology Platform for High-Performance Computing (ETP4HPC) Association and the Big Data Value Association.

**EuroHPC JU** is developing a world-class supercomputing Infrastructure and has started with the procurement and deployment in the EU three pre-exascale supercomputers (capable of at least 1017 calculations per second) and five petascale supercomputers (capable of at least 1015 calculations per second) that will be located across the European Union and will be available to Europe's private and public users, scientific and industrial users everywhere in Europe.

The three pre-exascale supercomputers will be located at the following supercomputing centres:

- Lumi (Large Unified Modern Infrastructure) in CSC – IT Center for Science, Finland
- LEONARDO (precursor to exascale supercomputers ) in CINECA, Italy
- Mare Nostrum 5 in Barcelona Supercomputing Centre, Spain

The five petascale supercomputers are located in the following centres:

- IZUM, Slovenia
- IT4Innovations National Supercomputing Center, Czech Republic
- Minho Advanced Computing Centre, Portugal
- Luxprovide, Luxembourg
- Sofiatech Park, Bulgaria

The **ESFRI Landmark PRACE** offers a pan-European supercomputing Infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level. PRACE aims to support all scientific disciplines that need HPC to achieve high impact discovery by offering world-class computing and data management resources and services through a centralised peer-review

process. PRACE members provide the computer systems and operations accessible through PRACE. Four hosting members – BSC representing Spain, CINECA representing Italy, GCS representing Germany and GENCI representing France – secured funding for the initial period from 2010 to 2016. PRACE has 26 members, representing European Union Member States and Associated Countries.

The PRACE RI has two forms of members:

- Members – a government organisation or legal entity representing a government. The PRACE RI accepts only one member per Member State of the European Union or an associated country as described in article 217 of the European Union Treaty. Further, to be eligible as a PRACE RI member the legal entity must be responsible for the provisioning of HPC resources and associated services.
- Hosting Members are members who have committed to fund and deliver PRACE RI computing and data management resources. There are 5 Hosting Members: France, Germany, Italy, Spain, and Switzerland.

In 2017, PRACE has engaged in the second period of the Partnership, securing the operations of the infrastructure until 2020, and adding a fifth Hosting Member, ETH Zurich representing Switzerland. During this second phase, PRACE will offer an initial performance above 62 Petaflops in 7 complementary leading-edge systems, offering a total of 4,000 million core-hours per year (75 million node hours).

PRACE also offers training services to users, through the PRACE Advanced Training Centre (PATC), PRACE Training Centres (PTC), PRACE seasonal school, and through online training material, including Massive Open Online Courses (MOOCs). Some joint training activities are provided by PRACE and EUDAT. PRACE is also using some services of GÉANT's network e-Infrastructure to provide European users access to Tier-0 systems. The PRACE project partners received funding from the EC under the PRACE Preparatory and Implementation Phase Projects for a total of € 97 million, complemented by the consortium budget of over € 60 million. PRACE is now in its 6th Implementation Phase Project. PRACE offers its computing services to projects or entities and the services of other e-Infrastructures (such as EOSC and GÉANT).

In terms of HTC, at the European level, **EGI** is a federated e-Infrastructure initially set up in order to provide advanced computing services for R&I using grid-computing techniques but which now also encompassed cloud computing Infrastructures. EGI is publicly funded and comprises over 300 data centres and cloud providers spread across Europe and worldwide. EGI offers a wide range of services for compute, storage, data and support. EGI has been funded by a series of EC research projects such as DataGrid and Enabling Grids for e-science.

EGI creates and delivers open solutions for science and RIs by federating digital capabilities, resources and expertise between communities and across national boundaries. Researchers from all disciplines have easy, integrated and open access to the advanced

scientific computing capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive science and innovation.

Regarding the services, EGI delivers advanced computing services to support scientists, multinational projects and RIs. The EGI services are provided by EGI's federated cloud providers and data centres. The services can be requested by anyone involved in academic research and businesses and they can be categorised in the following groups: computing, storage and data, training. EGI provides access to over 700,000 logical CPUs and 500 PB of disk and tape storage. The services can be requested by anyone involved in academic research and businesses and they can be categorised in the following groups: computing, storage and data, training. EGI provides access to over 700 000 logical CPUs and 500 PB of disk and tape storage.

---

## THEMATIC e-INFRASTRUCTURES

RIs are key elements of modern research. By providing services to a very broad variety of users, they create a shared and collaborative research environment, the so-called RI ecosystem, which has shaped big science for decades. In Europe, this includes the creation of the European Organization for Nuclear Research (CERN) in the mid-1950s, for particle physics research, and the European Southern Observatory (ESO) for astronomy in the early 1960s. From their early beginnings, both of these large RIs faced the challenge of managing large amounts of data they produced by developing data technologies and related policies.

RIs also had to develop schemes and processes to overcome challenges raised by the growth of the number of transnational RIs, the increased complexity of scientific problems and societal challenges (often requiring the collaboration of diverse user communities) and the exponential growth of data. Data protocols, quality control and management plans throughout the entire data lifecycle were developed along with the relevant technologies. Thematic RIs are therefore an indispensable and even a driving element of the EOSC data management chain.

The importance of thematic services provided to users of an RI and their interoperation with generic e-Infrastructures has been recognized by ESFRI by adding explicit attention to the development e-Needs in the lifecycle analysis of RIs. The ESFRI Roadmap thereby identifies the needs of the European scientific community in terms of Research Infrastructures including e-Infrastructures.

An example of this interdependency is the collaboration between CERN, SKAO, GÉANT and PRACE, which will see the organisations work together to help realise the full potential of the coming new generation of HPC technology. During an initial period of 18 months, the collaboration will develop a benchmarking test suite and a series of common pilot 'demonstrator' systems. The next-generation of HPC technology offers great promise for supporting scientific research. Exascale supercomputers – machines capable of performing a quintillion, or a billion billion, calculations per second – are

expected to become a reality in the next few years. This change in the power of HPC technology, coupled with growing use of machine learning, will be vital in ensuring the success of big science projects scheduled to come online this decade, such as SKAO and CERN's High-Luminosity Large Hadron Collider.

## GAPS, CHALLENGES AND FUTURE NEEDS

Despite the tremendous progression on Digital Research Infrastructures in general and e-Infrastructures for research in particular there is still a need for a (distributed) Research Infrastructure, which enables top quality computer science on the development of digital infrastructures itself. Another interesting development related to the collaboration between industry and publicly funded e-Infrastructures concerns GAIA-X<sup>10</sup>. This is a project initiated by Germany and France aiming to develop common requirements for a European Data Infrastructure. In September 2020, 22 companies and organisations – 11 from Germany and 11 from France as founding members – established an international non-profit association under Belgian law, the GAIA-X, European Association for Data and Cloud. The focus of GAIA-X is more industry oriented while EOSC is largely composed by public research organisations. The current collaboration opportunities and expectations between EOSC and GAIA-X are still unclear, although there are potential commonalities: both intend to create “a federated data infrastructure based on European values”. Common ground could emerge between GAIA-X and the next iteration of EOSC (serving industry) as described in the *Fair Lady report*.

The biggest challenges for the coming years will lie in further developing the EOSC concept into a working ecosystem, that serves the needs of the European research communities. These challenges broadly concern two levels: i) the interplay between generic and thematic e-Infrastructure service provisioning, and ii) the federation of institutional and national services up to the European level. Most of these challenges are organisational rather than technical.

The initial steps should lead to what is called a *Minimum Viable EOSC*, consisting of an EOSC Core, which should provide the functionality that is required to enable open science practices to occur across domains and countries according to the EOSC interoperability framework, the federated data and the EOSC Exchange, a digital marketplace that builds on the EOSC-Core to offer a progressively growing set of services exploiting FAIR data and encouraging its reuse by publicly funded researchers.

Many aspects need to be addressed among which the most important ones are: (i) implementation of this EOSC Core; (ii) definition of an architecture (Hardware and Software) that will allow to deploy new services; (iii) provide support and competences for turning to FAIR those data that are currently not compliant but are of interest for the scientific communities; (iv) create incentives for the researchers in making research data open and contribute to the creation of skills and careers in Data stewardship; (v) refine the sustainability and funding of EOSC with the correct mix of community support and business model.

10.

GAIA-X: A Federated Data Infrastructure for Europe  
<https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html>