

# Compute, storage and networking

## Introduction, context and scope

What are **e-Infrastructures**? The e-Infrastructure Reflection Group (e-IRG) [eIRG White Paper 2013], defines them to include: access to high-performance computing and high-throughput computing; access to high end storage for ever increasing data sets; advanced networking services to connect computing and storage resources to users and instruments; middleware components to enable the seamless use of the above services, including authentication and authorisation; and generic services for research, providing support for research workflows using combinations of the above (sometimes called virtual laboratories or virtual research environments). In particular, it envisions e-Infrastructures where the principles of global collaboration and shared resources are intended to encompass the sharing needs of all research activities.

The European Strategy Forum on Research Infrastructures (ESFRI) presented the European roadmap[1] for new, large-scale Research Infrastructures. These are modelled as layered hardware and software systems that support sharing of a wide spectrum of resources, spanning from instruments and observations, through networks, storage, computing resources, and system-level middleware software, to structured information within collections, archives, and databases. The roadmap recognises that the special “e-needs” of research infrastructures should be met by eInfrastructures.

Environmental and Earth sciences have been supported by national and institutional investments for a great many years. These have led to a diversity of significant computing resources and support services that are the precursors of today’s pan-European eInfrastructures. They now coexist with and participate in today’s pan-European eInfrastructures.

The contemporary supported strategies lead to the development of e-Infrastructures in Europe, connecting them into continent-wide e-Infrastructures. This is to allow researchers from different countries to work together using shared resources, including computers, data and storage. Important pan-European large-scale e-Infrastructures include: EGI, EUDAT, PRACE, GÉANT, OpenAIRE, and Helix Nebula. Each has own focused areas, e.g., EGI provides pan-European federated computing and storage resources; PRACE federates pan-European High Performance Computing (HPC) resources; EUDAT focuses on providing services and technology to support the life-cycle of data. GÉANT is the pan-European data network for the research and education community, interconnecting National Research and Education Networks (NRENs) across Europe. OpenAIRE is a network of Open Access repositories, archives and journals that support Open Access policies. 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. These e-Infrastructures provide generic IT resources and services solutions to support multiple European scientific research activities. The benefits to adopt and make good use of these resources for a scientific community and a research infrastructure include:

- Having ready-to-use compute and storage resources and services solutions for scientific collaborations;
- Avoiding duplicated development and effort;
- Enlarged community network and user bases – since these pan-European eInfrastructures have already been attracting many international collaborations and users;
- Sharing state-of-art experience by research communities already using the eInfrastructure.

This section gives an overview of current e-infrastructure for European research, along with some of the forthcoming developments and innovations. The focus is on pan-European scale infrastructure broadly classified into high-throughput computing (HTC or “cloud”; e.g., EGI), high-performance computing (HPC; e.g., PRACE), open-access publications repositories and catalogues (Pubs; e.g., OpenAIRE) and data storage and services (Data; the EUDAT CDI). The figure also includes a social dimension, characterising interactions by expert groups. The focus reflects the pan-European scale of the Research Infrastructures (RI) represented in ENVRIplus.

Figure 14: Classifying European e-Infrastructures

In general, all of the current European scale e-infrastructure seek to include partners in all European Member States, thereby providing a one-stop-shop for continental-scale interactions while at the same time providing access to local and regional activities in the individual Member States. At a European level, the e-infrastructure is often presented in terms of:

1. Computer networking
2. High capacity computing and high-throughput computing
3. Data storage and management
4. User tools (Virtual Research Communities and Virtual Research Environments).

In the sections 3.11.2 - 3.11.6 that follow we focus on the first three of these.

[1] ESFRI Roadmap: [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)

## Change history and amendment procedure

The review of this topic will be organised by in consultation with the following volunteers: . They will partition the exploration and gathering of information and collaborate on the analysis and formulation of the initial report. Record details of the major steps in the change history table below. For further details of the complete procedure see item 4 on the [Getting Started](#) page.

Note: Do not record editorial / typographical changes. Only record significant changes of content.

Date	Name	Institution	Nature of the information added / changed

## Sources of information used

The technology information is provided by e-Infrastructure providers, including EGI.eu, and CSC (representative of EUDAT). Information also refers to ESFRI Strategy Report on Research Infrastructure Roadmap 2016 [ESFRI 2016].

[ESFRI 2016] ESFRI, “European Strategy Report on Research Infrastructures: Roadmap 2016”. ISBN: 978-0-9574402-4-1, Mar 2016.

# Short term analysis of state of the art and trends

## Networking

### *GÉANT*

The model for research and education networking in Europe is of a single national entity per country (the National Research and Education Network – NREN) connecting to a common pan-European backbone infrastructure. In combination these networks provide a powerful tool for international collaborative research projects – particularly those with demanding data transport requirements. NRENs are able to connect individual sites to their high-bandwidth infrastructures or arrange point-to-point services for bilateral collaborations. GÉANT provides a single point of contact to coordinate the design, implementation and management of network solutions across the NREN and GÉANT domains.

The GÉANT network (like the majority of NRENs) has a hybrid structure – operating a dark-fibre network and transmission equipment wherever possible and leasing wavelengths from local suppliers in more challenging regions. This structure allows the operation of both IP and point-to-point services on a common footprint. Since 2013, GÉANT has migrated to a new generation of both transmission and routing equipment platforms. The resulting network is seen as a significant increase in the bandwidth available along with an improved range of network services. GÉANT's pre-provisioned capacity on each of the core network trunks (covering western and central Europe) is around 500Gbps and an advanced routing/switching platform delivers IP, VPN and point-to-point services with greater flexibility to all European NRENs.

The GÉANT project provides more than just a physical network infrastructure. Its service development and research activities address directly the needs of the R&E community both by providing advanced international services on the NREN and GÉANT backbones, and also by developing software and middleware to target network-related issues from campus to global environments. The GÉANT backbone currently offers:

- GÉANT IP – a high quality IP service providing robustness and high levels of availability, high-bandwidth and global reach.
- GÉANT Plus – point-to-point services offering guaranteed routing, latency and stability on the full GÉANT footprint.
- GÉANT Lambda – offering guaranteed capacity of 10Gbps or 100Gbps on dedicated wavelengths over the GÉANT-operated optical fibre.
- VPN (Virtual Private Network) services, which can provide bespoke network architectures for multi-site collaborations.

Services under development in GÉANT include[1]:

- Software-defined networking to facilitate faster and easier network configuration.
- Authentication and Authorisation (AAI) services – designed to address international multi-domain environments.
- A centrally procured cloud service to leverage economies of scale across the European NREN constituency.

GÉANT operates an infrastructure connecting NRENs in the vast majority of countries across Europe. These NRENs each have extensive national infrastructure and provide connections to universities, research centres and other not-for-profit institutions.

Seven new NRENs have joined GÉANT in 2013 from Eastern Europe and will be working to improve their international interconnection[2].

In addition to its pan-European reach, the GÉANT network has extensive links to networks in other world regions including North America, Latin America, the Caribbean, North Africa and the Middle East, Southern and Eastern Africa, the South Caucasus, Central Asia and the Asia-Pacific Region. In addition, there is on-going work to connect to Western and Central Africa[3].

## Computing

### *PRACE*

PRACE[4] provides high-end computing resources to European top science. The largest 3-5 PRACE systems are generally referred to as “tier-0” These systems are in general significantly larger than other European computer systems accessible to researchers. The resources are accessible to applicants with successful proposals submitted in response to Calls for Proposals. The "Guide for Applicants to Tier-0 Resources" on the PRACE website (<http://www.prace-ri.eu/HPC-access>) provides detailed information on preparing applications and the peer review process that follows the submission. Post Award obligations include a final report and acknowledgement of PRACE support. PRACE publishes twice-yearly Calls for Proposals, in February and in September. Preparatory access proposals, allowing users to develop software or test out novel ideas, are accepted at any time, with access granted on a quarterly basis.

The first phase of PRACE ended in mid 2015. PRACE now is in the second phase during which prototypes for the three most promising solutions will be built. Phase three is expected to start in early 2016 during which pre-commercial small scale product will be developed.

In addition to providing access to very large Tier-0 HPC resources, PRACE also pools some national level (Tier-1) resources and makes them available through specific calls. PRACE implementation projects include a range of activities that are likely to be interesting for the biological and medical community: training courses, software development, technology tracking, and access to prototype resources.

PRACE implementation projects include a range of activities that are likely to be interesting for the biological and medical science communities: training courses, software development, technology tracking and access to prototype resources. Three implementation projects have already been carried out (PRACE 1IP-3IP) and the fourth (PRACE 4IP) was funded in March 2015. PRACE 4IP aims to contribute to the biomedical application development, training needs and data intensive computing requirements, to name a few examples.

It is important to note that the explosion in the data generation capacity of scientific equipment and sensors is creating a new class of researchers who have different demands in terms of their use of computing power and of how and where their data is stored. Traditionally, users needed PRACE to develop tools to generate data, for modelling and simulations, which had to be kept to compare with other models. In contrast, the new type of users wants to analyse data generated elsewhere and tends not to have a strong background in computing. It is important to understand these users' requirements, in particular concerning how the data will be used, preserved and stored in the long-term.

### *EGI*

The EGI infrastructure is a publicly funded e-infrastructure put together to give scientists access to more than 650,000 logical CPUs, 550 PB of storage capacity to drive research and innovation in Europe. Resources are provided by about 350 resource centres who are distributed across 53 countries in Europe, the Asia-Pacific region, Canada and Latin America. EGI also federates publicly funded cloud providers across Europe for the implementation of an European data cloud to support open science.

EGI supports computing (including closely coupled parallel computing normally associated with HPC), compute workload management services, data access and transfer, data catalogues, storage resource management, and other core services such as user authentication, authorisation and information discovery that enable other activities to flourish. Resources are provided by over 350 resource centres that are distributed across 52 countries in Europe, the Asia-Pacific region, Canada, and Latin America. User communities gain access to EGI services by partnering with EGI, either directly through federating their own resource centres, or indirectly by accessing national or regional resource centres that already support their communities.

Existing high-level services:

- Federated IaaS Cloud: Run compute- or data-intensive tasks and host online services in virtual machines or docker containers on IT resources accessible via a uniform interface. Store/retrieve research data at multiple distributed storage service providers. Share applications, tools and software for data processing and analysis.
- High-Throughput Data Analysis: Run compute-intensive tasks for producing and analysing large datasets and store/retrieve research data efficiently across multiple service providers.
- Federated access to computing and data: Manage service access and operations from heterogeneous distributed infrastructures and integrate resources from multiple independent providers with technologies, processes and expertise offered by EGI.
- Consultancy for user-driven innovation: Expertise to assess research computing needs and provide tailored solutions for advanced computing.

High-level services under development:

- Open Data Platform: Store and discover research data, publish with open or controlled access, access and reuse data with the EGI computing services
- Accelerated computing: Run computational tasks on specialised processors (accelerators) with traditional CPUs from multiple providers allowing for faster real-world execution times.
- Community-specific tools: To provide access to specialised tools for data analysis contributed by the community

Project positioning with respect to similar initiatives

- EUDAT2020: EGI enables the reuse of research data available from their services
- PRACE: EGI complements their HPC services with cloud and HTC capabilities, altogether addressing the different computing needs of the research community
- GÉANT: EGI relies on their connectivity for distributed access to data and computing
- OpenAIRE: use of dissemination/discovery services for research outputs supported by EGI
- VRE projects: EGI provides hosting environments for services they are developing and we co-create community specific tools
- On-going project such as, INDIGO-DataCloud and AARC: EGI adopt their software and technical solutions

EGI matured its portfolio of solutions that help accelerate data-intensive research. The most relevant developments in EGI for ENVIplus are:

### ***1. Launch of EGI Federated Cloud***

After nearly two years of development the EGI community opened the 'EGI Federated Cloud' as a production infrastructure in May 2014. The new infrastructure (<http://go.egi.eu/cloud>) is based on open standards and offers unprecedented versatility and cloud services tailored for European researchers. It is a connected grid of institutional clouds built around open standards. With the EGI Federated Cloud, researchers and research communities can:

- Deploy scientific applications and tools onto remote servers (in the form of Virtual Machine images)
- Store files, complete file systems or databases on remote servers
- Use compute and storage resources elastically based on dynamic needs (scale up and down on-demand)
- Immediately address workloads interactively (no more waiting time as with grid batch jobs)
- Access resource capacity in 19 institutional clouds (the number is growing, see up to date values at [https://wiki.egi.eu/wiki/Fedcloudtf:ResourceProviders#Fully\\_integrated\\_Resource\\_Providers](https://wiki.egi.eu/wiki/Fedcloudtf:ResourceProviders#Fully_integrated_Resource_Providers))
- Connect their own clouds into a European network to integrate and share capacity, or build their own federated cloud with the open standards and technologies used by the EGI Federated Cloud.

Since its launch, the EGI Federated Cloud has attracted more than 35 use cases<sup>10</sup> from various scientific projects, research teams and communities. Among these there are several applications from environmental sciences:

### ***2. Simplifying access to EGI for the 'long tail of science'***

While processes to gain access to EGI are well established across the NGIs (National Grid Initiatives) for entire user communities, individual researchers and small research teams sometimes struggle to access compute and storage resources from the network of NGIs for the implementation of 'big data applications'. Recognising the need for simpler and harmonised access for individual researchers and small research groups, i.e. the 'long tail of science', the EGI community started to design and prototype a new platform in October 2014. The platform will provide integrated services from the NGIs to those researchers and small research teams who work with large data but have limited or no expertise in using distributed systems. The platform will lower the barrier to access grid and cloud infrastructure via a centrally operated access management portal and an open set of virtual research environments designed for the most frequent use cases. The project defines security policies and implements new security services that enable personalised, secure and yet simple access to e-infrastructure resources via the virtual research environments for individual users. The platform will authenticate users via the EduGAIN federation and other username–password based mechanisms, complementing the long established certificate-based access mechanisms. The prototype system is launched in Dec 2015. (<https://access.egi.eu/start>)

### ***3. End of EGI-InSPIRE, start of EGI-Engage***

EGI's first nearly 5 years were supported by the 'EGI-Integrated Sustainable Pan-European Infrastructure for Research in Europe' (EGI-InSPIRE) FP7 project. EGI-InSPIRE came to an end in December 2014. A new initiative, EGI-Engage was funded by the European Commission for support under the H2020 framework programme. EGI-Engage was launch in March 2015 with a total budget of 8.7 million Euros for 2.5 years.

One of the main objective of EGI-Engage is to expand the capabilities of EGI (e.g. cloud and data services) and the spectrum of its user base by engaging with large Research Infrastructures (RIs), the long tail of science, and with industry/SMEs (Small and medium-sized enterprises). The key engagement instrument for this is a network of eight Competence Centres, in which National Grid Initiatives (NGIs), user communities, technology and service providers are join-forces to collect requirements, integrate community-specific applications into state-of-the-art services, foster interoperability across e-infrastructures, and evolve services through a user-centric development model. The competence centres will provide state-of-the-art services, training, technical user support and application co-development to specific scientific domains. The following science communities have dedicated Competence Centres in EGI-Engage:

1. Earth-science research (EPOS)
2. EISCAT 3D
3. Life-science research (ELIXIR)
4. Biodiversity and ecosystem research (LifeWatch)
5. Biobanking and medical research (Biobanking and Biomolecular Research Infrastructure, BBMRI-ERIC),
6. Structural biology and brain imaging research (MoBrain supporting WeNMR and Integrating Structural Biology – INSTRUCT)
7. Arts and Humanity (DARIAH)
8. DisasterMitigation

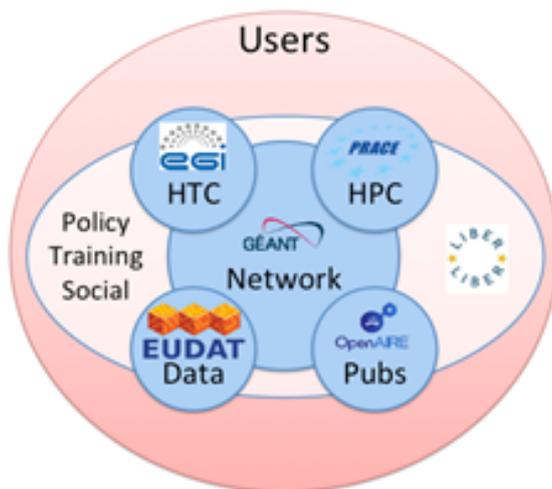
### *The Helix Nebula Marketplace*

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 European Commission (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. Also, GÉANT is actively helping NRENs (National Research and Education Networks) to deliver cloud services to their communities. It is engaging with the existing NREN brokerages to promote an efficient and coordinated panEuropean approach, by building on existing experience and supplier relationships [ESFRI 2016].

### *EUDAT*

EUDAT is a pan-European data infrastructure initiative. EUDAT brings together a large consortium of 33 partners, including research communities, national data and high performance computing (HPC) centres, technology providers, and funding agencies from 14 countries. EUDAT aims to build a sustainable cross-disciplinary and cross-national data infrastructure that provides a set of shared services for accessing and preserving research data.

The EUDAT *Collaborative Data Infrastructure* (CDI) is a defined data model and a set of technical standards and policies adopted by European research data centres and community data repositories to create a single European e-infrastructure of interoperable data services. The EUDAT CDI is realised through ongoing collaboration between *service providers* and *research communities* working as part of a common framework for developing and operating an interoperable layer of common data services. The scope of the CDI covers data management functions and policies for upload and retrieval, identification and description, movement, replication and data integrity. EUDAT's vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres. At the heart of the CDI is a network of distributed storage systems hosted at the major scientific data centres. Between them, these centres manage more than 100 PB of high performance, online disk in support of European research, plus an even greater amount of near-line tape storage. EUDAT's strength lies in the connections between these centres, the resilience resulting from the geographically distributed network, and its ability to store research data right alongside some of the most powerful supercomputers in Europe.



In defining the EUDAT CDI's position with respect to other e-infrastructure initiatives and organisations, EUDAT regards any and all e-infrastructures (including, though not limited to, PRACE, EGI, HelixNebula, OpenAIRE) as *organisational end-users* of EUDAT's services. The CDI Gateway API defines a clear contract with external end-users and consequently a set of stable targets for computational jobs (scripts, programs or workflows) running on external infrastructure..

We can classify the European e-infrastructures, broadly, into high-throughput computing (HTC or "cloud"; e.g. EGI), high-performance computing (HPC; PRACE), open-access publications repositories and catalogues (Pubs; e.g. OpenAIRE) and data storage and services (Data; the EUDAT CDI). We include a social dimension, characterizing interactions by expert groups.

The key value that EUDAT's implementation of the CDI brings to any external user is a *well defined API to EUDAT services* and *coherent service offerings* across all EUDAT partner sites. These common, coherent service interfaces create the line of demarcation between the EUDAT CDI and the other e-Infrastructures – the boundary of the domain of registered data. Other infrastructures then have clear ways to interact with the EUDAT CDI. Across the *network* they can:

- retrieve metadata records by PID (e.g. HTC workflows, HPC programs, publication repositories & catalogues);
- retrieve open access data by PID (e.g. HTC workflows, HPC programs, publication repositories & catalogues);
- subscribe to metadata feeds using OAI-PMH (e.g. publication catalogues);
- where authorised: create (upload) data & metadata and receive a registered PID (e.g. HTC workflows, HPC programs & scripts);
- where authorised: update or delete data and/or metadata by PID (e.g. HTC workflows, HPC scripts).

This model positions the EUDAT CDI clearly as the home for persistent, shared, re-used research data.

EUDAT is about preserving research data for reuse, and an aspect of making digital data reusable lies in providing the capabilities for efficient computation on them. EUDAT2020 enables data analytics by staging data to dedicated analysis systems – leveraging the computing capacity made available via EGI and PRACE. EUDAT has issued two joint public calls in 2015 with PRACE allowing PRACE users which have been granted PRACE computing resources to store the data resulting from simulations into EUDAT. It is also working with EGI to strengthen interoperability between the two infrastructures with a view to connect data stored in the EUDAT Collaborative Data Infrastructure to high throughput and cloud computing resources provided by EGI. EUDAT develops solutions for data coupled computing, including big data frameworks and workflow systems for initiating computing tasks on datasets located in the EUDAT infrastructure. EUDAT B2STAGE library allows to stage data to HPC computing environments and is being developed further to add support for Hadoop and Spark big data systems. EUDAT also offers a hosting environment for the deployment and provision of data analytics services directly at the data centres – building on the Service Hosting Framework successfully trialled in the first EUDAT project to provide a flexible virtual computing environment at participating data centres, a highly-configurable cluster computing platform sited right alongside the data archives.

## Data

### Research Data & EUDAT

Currently, EUDAT is working with more than 30 research communities covering a wide range of scientific disciplines and has built a suite of integrated services to assist them in resolving their technical and scientific challenges.

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.

Table 1: The EUDAT Service Catalogue

Service	Function	Status	Individual Researcher	RI/ Community Manager	Service Provider
<b>Data Discovery</b>					
<b>B2FIND</b>	Multi-disciplinary joint MD catalogue	Active	X	X	
<b>Metadata Catalogue</b>	MD extraction, MD store, index	Under develop.		X	X
<b>Data Hosting, Registration &amp; Management &amp; Sharing</b>					
<b>B2DROP</b>	Cloud storage, sync & exchange	Active	X	X	X
<b>B2SAFE</b>	Policy-driven data management	Active		X	X
<b>B2SHARE</b>	Repository for sharable digital objects	Active	X	X	X
<b>B2HANDLE</b>	Policy-based prefix & PID management	Active		X	X
<b>Data Type Registry</b>		Under develop.		X	
<b>Data Access, Interface &amp; Movement</b>					
<b>B2ACCESS</b>	Federated multi-protocol IAM	Active	X	X	X
<b>Generic API</b>	Common data interface service	Under develop.	X	X	X
<b>B2STAGE</b>	Data staging service CDI ext.	Active	X	X	X
<b>Subscription</b>	Data transfer subscription	Under develop.		X	X
<b>Consultancy</b>					
<b>Training</b>	on services & data management	Active	X	X	X
<b>Consultancy</b>	on licensing, certification, data privacy, data system design	Active	X	X	X
<b>Helpdesk</b>	Support and enabling	Active	X	X	X
<b>Operations</b>					
<b>Service Hosting</b>	PaaS, IaaS, SaaS	Under develop.		X	X
<b>Monitoring</b>	Availability & reliability monitoring	Active		X	X

<b>Accounting</b>	Storage & Data Usage Reporting	<b>Under develop.</b>		<b>X</b>	<b>X</b>
<b>SLC Management</b>	Service Portfolio & Catalogue	<b>Active</b>	<b>X</b>	<b>X</b>	<b>X</b>
<b>Coordination</b>	Project Implementation, Service & Resource Provisioning	<b>Active</b>	<b>(X)</b>	<b>X</b>	<b>X</b>
<b>Site Registry</b>	Site, Service & Service Groups	<b>Active</b>	<b>(X)</b>	<b>X</b>	<b>X</b>

These services have been developed together with research communities coming mostly from the environmental sciences (EPOS, ICOS, EISCAT, ENES, LTER, DRIHM), life sciences (ELIXIR, VPH, BBMRI, ECRIN, DIXA), and social sciences and humanities (CLARIN, CESSDA, DARIAH). In October 2015, EUDAT issued a public call for data pilot and received 24 applications including 9 from Earth and Environmental Sciences, Energy and Environment disciplines, 6 from the Biomedical and Life Sciences, 5 from the Social Sciences and Humanities, and 4 from Physical Sciences and Engineering. Altogether these pilots represent a potential user base of 40,000 researchers.[5]

EUDAT distinguishes three main types of users/customers of its services and infrastructure:

- Individual researchers: Those wishing to share data with colleagues or collaborators, or those wishing to discover and re-use data as part of their ongoing research.
- These users are anybody – researchers (from academia and industry), citizen scientists, policy makers, and members of the public – anyone wanting to share or re-use European research data in simple, powerful ways.
- As a user your main responsibility is to adhere to the terms and conditions of the B2 services provided by the EUDAT consortium
  - Organized research communities: Those concerned with the management of their research e-infrastructure and/or community-specific data repositories who wish to join their repositories formally with the CDI network or deploy EUDAT services on top
  - These research communities are organized research groups (e.g. EC projects), research Infrastructures (e.g. ESFRI) or universities and libraries – anyone interested in archiving, replicating, processing and cataloguing data on behalf of the research community they support.
  - They either use EUDAT services a) as they are, according to the service terms and conditions, b) through an agreement with a specific service provider, or c) by joining the CDI as a node/service provider.
    - CDI Service providers
    - Service Providers wish to use and/or deploy CDI services to support or augment their existing role and service portfolio – to provide long term preservation of important digital assets, offer wider accessibility, intelligent caching of data near compute, data integrity checking and so on.

### *Research Data Alliance*

Together, EUDAT and OpenAIRE are driving international cooperation in tackling issues around large-scale data infrastructures through the recently formed international Research Data Alliance (RDA). The RDA is an international collaboration including participants from all around the world. In addition to EUDAT and OpenAIRE, the EC and NSF are directly represented in RDA. In Europe, the work of the RDA has been supported by a series of projects under FP7 and H2020.. The RDA aims to accelerate and facilitate research data sharing and exchange. The work of the RDA is primarily undertaken through its working groups. Participation in working groups, starting new working groups, and attendance at the twice-yearly plenary meetings is open to all.

### *Open Data Commons of EGI*

EGI developed its 'Open Data Commons' vision inspired by the emerging open access policy in the European Research Area. The goal of open access is to ensure that research results are made available free of charge to end users and that are reusable. Research results thus become a shared community resource (i.e., a commons). In order for this to happen, researchers need to change their own behaviours and they need to be supported with services that simplify the sharing of research results, their discovery and reuse. In the EGI-Engage project (starting in March 2015) EGI will develop the concept of a federated open research data platform, an innovative solution enabling to publish data, link to open access repositories, and offering easy integration into processing capabilities (e.g. EGI Federated Cloud). Furthermore, the federated cloud infrastructure, including existing publicly funded institutional cloud and expanding to commercial clouds, will evolve to offer IaaS, PaaS and SaaS for specific communities, the long-tail of research and the industrial/SME sector. In collaboration with other e-infrastructure, services will be tailored to meet the needs of the long tail of research and their evolution will be driven by the requirements of the RIs on the ESFRI roadmap that participate in the EGI Engage project through Competence Centres.

### *Publications & OpenAIRE*

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 institutional or a subject repository, Zenodo, 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 (European Research Council). 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 [ESFRI 2016].

[1] For full details of GÉANT services see <http://www.geant.org/Services>.

[2] GÉANT NRENs list: <http://www.geant.org/About/Membership/Pages/MAandGareps.aspx>

[3] Countries interconnect with GÉANT [http://www.geant.org/Networks/Global\\_networking/Pages/Home.aspx](http://www.geant.org/Networks/Global_networking/Pages/Home.aspx).

[4] PRACE: <http://www.prace-ri.eu>

[5] <https://www.eudat.eu/eudat-communities-pilots>

[ESFRI 2016] ESFRI, "European Strategy Report on Research Infrastructures: Roadmap 2016". ISBN: 978-0-9574402-4-1, Mar 2016.

## A longer-term horizon

In the recent published[1]ESFRI Roadmap 2016, it highlights the notion of a *European e-infrastructure Commons* referring to the framework for an easy and cost-effective shared use of distributed electronic resources for research and innovation across Europe and beyond. The concept is outlined by the e-Infrastructure Reflection Group (e-IRG) based on the identification of the need for a more coherent e-infrastructure landscape in Europe.

According to the e-IRG report[2],

*“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 infrastructure communities.
- Provisioning of sustainable and interoperable e-infra structure 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 on the European level and globally. This e-infrastructure Commons is also a solid basis for building the *European Open Science Cloud* as introduced in the description of the Digital Single Market[3], already containing most of the ingredients needed for an integrated European platform for Open Science [ESFRI 2016].

To support this vision, it would request a long-term agenda for supporting a coherent, innovative and strategic European e-infrastructure policy making and the development of convergent and sustainable e-infrastructure services. Today (April 2106) the EC announces the European Cloud Initiative[1] - €6.7billion of public and private investment in European Open Science Cloud (2016), opening up by default all scientific data (2017), flagship initiative on quantum technology (2018), development and deployment of European high performance computing, data storage and network infrastructure (2020), including by acquiring two prototype next-generation supercomputers of which one would rank among the top three in the world, establishing a European big data centre, and upgrading the backbone network for research and innovation (GEANT).

---

[1] [http://europa.eu/rapid/press-release\\_IP-16-1408\\_en.htm?locale=en](http://europa.eu/rapid/press-release_IP-16-1408_en.htm?locale=en)

---

[1] [ESFRI 2016] ESFRI, “European Strategy Report on Research Infrastructures: Roadmap 2016”. ISBN: 978-0-9574402-4-1, Mar 2016.

[2] <http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf>

[3] SWD(2015) 100 final accompanying the document “A Digital Single Market Strategy for Europe” COM(2015) 192 final, SWD(2015) 100 final

## Relationships with requirements and use cases

ENVRiplus has already been collaborating with these pan-European e-Infrastructures, such as EGI and EUDAT. EUDAT services are chosen (by some of Research Infrastructures) for data management.

In ENVRiplusWP9, EGI will provide computing and storage resources for deploying services developed by ENVRiplus development WPs. The task begins with identifying a number of community use cases, and the feasibility of deployments of the use cases are evaluated by e-Infrastructure experts. 5-6 use cases are selected which will have resources and technical supports from EGI for deployments.

## Summary of analysis highlighting implications and issues

Interoperable access to these e-Infrastructures remains as a challenging issue. In this sense, ENVRiplus is in good position to provide real use cases/requirements to influence the future implementations of these e-Infrastructures.

## Bibliography and references to sources

[ESFRI 2016] ESFRI, “European Strategy Report on Research Infrastructures: Roadmap 2016”. ISBN: 978-0-9574402-4-1, Mar 2016.